

Choice-Based Revenue Management: A Hotel Perspective

A Dissertation

Presented to

The Academic Faculty

by

Tudor Dan Bodea

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy in the

School of Civil and Environmental Engineering

Georgia Institute of Technology

August 2008

COPYRIGHT 2008 BY Tudor Dan Bodea

Choice-Based Revenue Management: A Hotel Perspective

Approved by:

Dr. Laurie Garrow, Advisor
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Marco Castillo
School of Public Policy
Georgia Institute of Technology

Dr. Patrick McCarthy
School of Economics
Georgia Institute of Technology

Dr. Mark Ferguson
College of Management
Georgia Institute of Technology

Dr. Michael Meyer
School of Civil and Environmental Engineering
Georgia Institute of Technology

Date Approved: April 2008

*To
Andreea KISS,
Ana and Todorut BODEA,
and
in memory of
Maria NEAG.*

ACKNOWLEDGEMENTS

I would like to thank my family for the support and understanding they showed me through my academic life at the Georgia Institute of Technology. My wife, Andreea Kiss, deserves special recognition since I would have never become the person that I am today without her being by my side.

I would like to express my gratitude to all members of my Ph.D. committee - Dr. Marco Castillo, Dr. Mark Ferguson, Dr. Laurie Garrow, Dr. Patrick McCarthy and Dr. Michael Meyer - for their continuous support and invaluable suggestions. I thank my advisor, Dr. Laurie Garrow, for her constructive advice and willingness to share her insight and wisdom. Dr Michael Meyer deserves my recognition for being an inspiring leader and a visionary mentor. I offer my sincere thanks to Dr. Mark Ferguson whose endless efforts to teach and guide me through Revenue Management kept me motivated.

My gratitude also goes to former and current students in various graduate programs at the Georgia Institute of Technology for their true friendship and unconditional support. With the risk of omitting some, my thanks go to Mr. Saroch Boonsiripant, Mr. Radu Cazan, Dr. Hsing-Chung Chu, Dr. John Crocker, Dr. Anne Dunning, Dr. Angshuman Guin, Mr. Petru Horvath, Mr. Dan Iliescu, Dr. Ionut Porumbel, Mr. Muthukumar Subrahmanyam and Dr. Danena Lewis Gaines.

Last but not least, I would also like to thank the hotel chain that sponsored this research study for the support, the guidance and the friendship they so generously extended to me. The Global Revenue Management and Pricing teams substantially contributed to the findings of this study.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iv
LIST OF TABLES	viii
LIST OF FIGURES	x
LIST OF SYMBOLS AND ABBREVIATIONS	xii
SUMMARY	xiii
Chapter 1. INTRODUCTION	1
1.1. Background and Motivation	1
1.2. Research Objectives	4
1.3. Major Contributions	6
1.4. Dissertation Structure	8
Chapter 2. BACKGROUND	10
2.1. Overview of Revenue Management Core Concepts	10
2.2. Introduction to Choice-Based Revenue Management	13
2.3. Choice-Based Revenue Management: Summary of the Relevant Literature	15
2.4. Modeling Customer Behavior: Single-Leg Revenue Management Problems	18
2.5. Modeling Customer Behavior: Network Revenue Management Problems	25
2.6. Summary	37
Chapter 3. CONCEPTUAL MODEL AND METHODOLOGY	40
3.1. Introduction	40
3.2. Conceptual Model - Abstract Representation	41
3.3. Conceptual Model - Details and Supporting Arguments	42
3.4. Modeling Customer Purchasing Behavior	47
3.4.1. Multinomial Logit Model	47
3.4.2. Expectation Maximization Algorithm	58
3.5. Computing Preferred Capacity Control Policies	69
3.5.1. Expected Marginal Seat Revenue Heuristic - Version b	69

3.5.2. Optimal Choice-Based Capacity Control Policy	76
3.6. System Revenue Performance Evaluation	82
3.7. Summary	88
Chapter 4. DATA	90
4.1. Introduction, Motivation and Purpose	90
4.2. Data Collection	92
4.2.1. Data Collection Methodology	92
4.2.2. Data Collection Plan - Measurement Capability	97
4.2.3. Descriptive Statistics	98
4.2.4. Data Description	103
4.2.5. Discussion of Data Strengths and Limitations	105
4.3. Summary	108
Chapter 5. RESULTS	110
5.1. Preface	110
5.2. Introduction and Motivation	117
5.3. Lowest Fare Available Model	119
5.4. Capacity Control Techniques	120
5.4.1. EMSR-b Heuristic	120
5.4.2. Choice-Based Capacity Control Technique	121
5.4.3. Multinomial Logit Model	124
5.4.4. Alternative Approach to Multinomial Logit Model	128
5.5. Experimental Setup	130
5.6. Capacity Control Policies	131
5.6.1. EMSR-b Heuristic	132
5.6.2. Choice-Based Capacity Control Technique	136
5.7. Revenue Performance Evaluation	142
5.8. Summary	148

Chapter 6. SUMMARY, FINDINGS AND RECOMMENDATIONS	150
6.1. Summary	150
6.2. Contributions and Findings.....	151
6.3. Limitations and Future Research	154
Appendix A. EMSR-b REVENUE PERFORMANCE EVALUATION - ALGORITHM DESCRIPTION.....	158
REFERENCES	165
VITA.....	173

LIST OF TABLES

Table 1 Summary of Relevant Choice-Based RM Literature	39
Table 2 MNL Model - Input Data.....	52
Table 3 MNL Model - Excerpt from Parameter Estimation (Intermediate Step).....	53
Table 4 MNL Model - Average Customer's Purchasing Probabilities (Example).....	56
Table 5 EM Algorithm - Problem Set Up.....	67
Table 6 EMSR-b - Input Data and Computed Protection Levels	73
Table 7 Optimal Control Policy for Model 1	82
Table 8 Optimal Control Policy for Model 2.....	82
Table 9 Latent Demand Distribution - Hypothetical Example	86
Table 10 Characteristics of the Study Hotels.....	93
Table 11 Product Availability and Display Order by Distribution Channel.....	95
Table 12 Consumed Bookings by Rate Type and Hotel for Study Check-In Dates.....	99
Table 13 Data Descriptive Statistics	102
Table 14 Description of Data Set Fields	104
Table 15 Revenue Performance Assessment - Methodology and Examples.....	114
Table 16 Parameter Estimates for the Unconstrained Product Demand Distributions...	135
Table 17 EMSR-b Heuristic - Nested Protection Levels	135
Table 18 MNL Parameter Estimates Computed using the EM Method	137
Table 19 MNL Choice Probabilities	137
Table 20 Parameter Estimates for the Unconstrained Product Demand Distributions: Demand Dependent upon the Offered Product Sets	139
Table 21 Choice Probabilities - Alternative Approach.....	140
Table 22 Optimal Total Expected Revenue $V_t(x)$	141
Table 23 Revenue Performance and other Performance Measures	142
Table 24 Revenue Performance Comparison - Two-Sample Paired t-test Results	144

Table 25 Study Limitations..... 157

LIST OF FIGURES

Figure 1 Optimal Control Policy - Nesting by Fare Order	23
Figure 2 PODS System Flow [Original in Hopperstad [55], p. 5; Adapted from Belobaba and Wilson [42], p. 5]	28
Figure 3 Customer Flight Selection - Preference Mapping & Seat Availability	31
Figure 4 Conceptual Model	42
Figure 5 Refinement of Conceptual Model	46
Figure 6 Customer Arriving Process - Example	59
Figure 7 EMSR-b - Graphic Summary of Embedded Assumptions	70
Figure 8 EMSR-b - Relationship between Protection Levels and Product Class Availability	74
Figure 9 Efficient Frontier and Efficient Sets	81
Figure 10 Revenue Performance Evaluation	88
Figure 11 Data Collection Process	93
Figure 12 Selling Process by Distribution Channel	94
Figure 13 Potential Sources of Biases	100
Figure 14 Data Set Format for Hotel 1, Bookings 44-47	104
Figure 15 Revenue Performance Comparison: EMSR-b Static vs. EMSR-b Dynamic .	115
Figure 16 Revenue Differences between Choice-based RM and EMSR-b Dynamic: Two-Sample 95 % Confidence Intervals	115
Figure 17 Revenue Performance Comparison: Competing Customer Purchasing Model Specification	116
Figure 18 Revenue Performance Comparison: Competing Imposed Business Rules	116
Figure 19 Box-Plots of the Booking Limits by Arrival Batch (Capacity 140; 50 Repetitions)	131
Figure 20 Incremental Product Demand Plots for a Hypothetical Arrival Stream	133

Figure 21 Unconstrained Product Demand Levels for a Hypothetical Arrival Stream ..	135
Figure 22 Unconstrained Product Demand Levels for a Hypothetical Arrival Stream: Demand Dependent upon the Offered Product Sets	138
Figure 23 Optimal Choice-Based Capacity Control Policy - Example	142
Figure 24 Box-Plots of Revenues Reported by Competing RM Methods.....	144
Figure 25 Revenue Comparison: EMSR-b vs. Ch-B vs. Ch-B EM.....	145
Figure 26 Sales by RM method and Product Price - Capacity 90 units.....	146
Figure 27 Sales by RM method and Product Price - Capacity 110 units.....	147
Figure 28 Sales by RM method and Product Price - Capacity 120 units.....	148

LIST OF SYMBOLS AND ABBREVIATIONS

List of Abbreviations - Hospitality Industry Terminology

CRO	Central Reservation Office - centralized offices where customers call directly to secure hotel reservations. All major lodging service providers have multiple CROs worldwide.
CRS	Central Reservation System - system that offers real-time rates and room availability information to all interested parties (e.g., hotel reservation staff, CRO's employees, on-line customers, etc.). Typically, each hotel operator owns its own CRS.
GDS	Global Distribution System - electronic travel reservation system available to travel agencies throughout the world. GDSs (e.g., Amadeus, Galileo, Sabre, Worldspain) are owned by external enterprises and are made available to all interested hotel operators.
LP	Loyalty Program - hotel rewards program design to incentivize travelers to book with the same hotel chain.
RM	Revenue Management - business concept that tries to sell the right product to the right customer through the right distribution channel at the right time and the right price in such a way that a firm's revenues are maximized.
SAS	Statistical Analysis System (SAS Institute Inc. [1]) - comprehensive collection of statistical routines.
TPI	Third Party Intermediaries - generic name given to all on-line hospitality service providers and room consolidators that are not affiliated with the hotel chain but sell its products (e.g., Travelocity, Orbitz, etc.).

SUMMARY

This study investigates the revenue performance of choice-based revenue management (RM) systems in various business environments. Previous research conducted using simulated data suggests that incremental revenue gains of up to 15% are to be expected when choice-based RM techniques are employed. In addition, despite the novelty of these techniques, the implementation of choice-based RM systems is considered to be feasible at large global corporations. The revenue potential and the ease of execution associated with the choice-based methods are examined in the context of a large hotel chain. Customer-centric data which includes transaction and time of booking availability information is collected for five hotel properties located in the continental US. The customer preference for hotel products and their attributes is determined using discrete choice and other ad hoc models of demand. Optimization techniques that account for the customer purchasing behavior are employed to compute the capacity control policies the hotel operator should follow to maximize its revenues. Results indicate that collecting customer-centric data from today's RM systems is a time-consuming task. In the environment in which the study hotels operate, the choice-based RM systems report incremental revenue gains that are dependent on how the purchasing behavior models are formulated. In capacity constrained regimes that are the focus of RM, revenue gains of up to 2% are typically noted. In controlled environments in which the customer purchasing behavior can be better asserted, the incremental revenue gains range between 1% and 14%. These findings suggest that the execution of the choice-based RM, while feasible, needs to be preceded by the implementation of efficient and, most likely, expensive data collection procedures. The incremental revenue gains, consistent with those reported in the literature, indicate that RM users can substantially benefit from the use of the choice-based RM.

Chapter 1. INTRODUCTION

There is nothing more difficult to take in hand, more perilous to conduct or more uncertain in its success than to take the lead in the introduction of a new order of things.

Niccolo Machiavelli, Italian writer and statesman
(www.thinkexist.com)

1.1. Background and Motivation

WAR. A word that inspired countless emotional stories. A word that puts tears and fears in people's eyes and soul everyday. A word that is associated with destruction, pain, and sufferance while also promoting fresh starts and freeing positive energies. A word that for a few lucky some may mean a better tomorrow. A word that sometimes opens new opportunities.

A war, although a business war, contributed fundamentally to shaping the concepts based on which the Revenue Management (RM) field, as it is known today, evolved. It was the bitter fight between two opposing business ideas that led to the inception of an innovative approach to improving profitability. It was the attempt of an entrepreneur to open "the skies to millions of new air travelers" (Cross [2], p.101) that forced a giant to employ all of its intellectual resources to regain and conquer the air markets the entrant threatened. It was the "Flying That Costs Less Than Driving" marketing slogan of the former that finally gave the latter the right weapons to fight and win this battle. It was the unfortunate decision of the entrant not to invest in information technology that provided the incumbent with the competitive advantage that resulted in

an aggressive fare war in all markets flown by the newcomer. And in all this, it was the chance which favored the decline of an otherwise revolutionary business idea.

The start of the war that transformed the airline industry was announced by the adoption of the Airline Deregulation Act (ADA) of 1978 whose declared purpose was “to encourage, develop, and attain an air transportation system which relies on competitive market forces to determine the quality, variety, and price of air service” (ADA [3]).

People Express, a new no-frills airline with reduced capital and maintenance costs and an energetic work force, was committed to do whatever it takes to “bring air transportation to the masses” (Cross [2], p. 101). The strategy adopted by People Express consisted in offering low fare tickets for destinations on which established carriers were already operating. The new strategy adopted resulted in a massive market share increase for People Express and added visibility to a player previously ignored by the mainstream competitors in the field. However, People Express’ competitive advantage was only temporary as its more technologically advanced competitor, American Airlines, gains its lost market shares back by increasing its ability to correctly forecast its empty seats and price them competitively. More specifically, the improved Yield Management system allowed American Airlines to sell its full-fare tickets while also guaranteeing the sale of any surplus tickets at prices below those used by low cost carriers such as People Express. The stage was thus set. The simple yet innovative system implemented by one of the major players in the industry became “best practice” for most of the major carriers in the US.

The approach initiated by American Airlines in the mid 1980s was closely analyzed and later improved by both industry practitioners and academic researchers. However, the improvements to this approach were incremental, and did not challenge its fundamental assumptions. The demand for a certain product fare class was assumed to be independent from that for other fare classes, and its estimation did not consider the customer’s characteristics or the product attributes. As a consequence, the price

transparency enabled by the Internet and the discounted fares offered by low cost carriers starting with the late 1990s strongly impacted the yield performance of the RM systems in use and rendered them obsolete.

The last years, however, have witnessed increased efforts to improve the performance of the traditional RM systems by making them more responsive to the changing competitive market environment. In this context, many practitioners and academic researchers have started to openly challenge these systems' fundamental assumptions and propose alternative and/or possible solutions (Boyd [4], Boyd and Kallesen [5], Dunleavy and Westermann [6], Hoang [7], Hornick [8], Lieberman [9], Nason [10], Oliveira [11], Ratliff [12], Sfodera [13], and Talluri and van Ryzin [14]). In particular, it is believed that incorporating customer purchasing behavior as well as the competition interdependencies into RM systems will help the field regain its lost credibility (van Ryzin [15]). To this end, the approaches proposed by Talluri and van Ryzin [14], Zhang and Cooper [16], Gallego, et al. [17], van Ryzin and Liu [18], van Ryzin and Vulcano [19], [20], and Bront, et al. [21] are closer to reality as they account for the customer purchasing behavior when recommending the preferred capacity control policies. In the same context, but taking into account the fact that the customer valuations for products may change over time, Gallego and Sahin [22] investigate the effectiveness of various pricing strategies by employing inter-temporal choice models. Hu and Gallego [23] link pricing and capacity allocation in the presence of competitive information using a demand model that also accounts for the consumer's choice.

The papers above report promising results associated with the inclusion of customer purchasing behavior into the basic capacity allocation policies of RM systems but the majority use simulation studies to make these claims with very limited testing on real data sets. The use of simulated data is understandable given the difficulty of collecting all the needed information in real settings, yet rigorous testing on real data is needed to convince a skeptical user community of the benefits associated with

incorporating consumer choice models into their RM systems¹. While insights into other related business areas can be explored, the focus of this work is to better understand how the customer purchasing behavior as modeled using real customer-centric data impacts the capacity control policies, and implicitly, the revenue performance at a major hospitality service provider.

The remainder of this chapter is organized as follows. Section 1.2 describes the research objectives of this work. Next, Section 1.3 details the major contributions of this study. Finally, Section 1.4 outlines the organization of the dissertation.

1.2. Research Objectives

This study has four interrelated research objectives that are listed in the order in which they are treated in the body of the study.

Objective 1: *Provide a better understanding of the choice-based RM system's data requirements.*

As mentioned in the introductory section, the development of a choice-based RM system requires customer-centric data. Customer-centric data forces service operators to complement transaction level, product oriented data with, at least, time of booking product availability information. To meet this requirement, a data collection methodology, which enabled a major hotel chain to compile several customer-centric data sets, was devised and, subsequently, administered. This study provides an in-depth look at the proposed data collection methodology and offers new insights into practical issues that need to be addressed to successfully implement choice-based RM systems in the hospitality and other similar industries.

¹ The difficulties associated with collecting and working with real customer data are acknowledged by many academic researchers and industry practitioners. Professor van Ryzin, referring to this problem, concluded that “the complexity and cost of dealing with such data [i.e., customer-centric data] is a daunting challenge in building customer-level models of demand” (van Ryzin [15]).

Objective 2: *Learn the extent to which the choice-based RM methodologies succeed in integrating the customer purchasing behavior into the corresponding RM framework.*

Many choice-based RM methodologies rely on a behavioral component to feed the standard optimization routines with required customer related inputs regardless of how the actual methodologies are set up. In most formulations, the behavioral component provides the downstream processes with information about customers' preference for products and/or product characteristics. This information, assumed to be accurate, assists in the search for the preferred capacity control policy. An attempt is then made to understand whether the transferred information is consistent with the expected customer purchasing behavior. In this context, the concern is that the algorithms devised to behaviorally describe customers' purchasing decisions may fail to do so under specific conditions.

Objective 3: *Understand the impact that various specifications of the customer purchasing behavior model have on downstream choice-based RM processes.*

The customer purchasing behavior models describe, in probabilistic terms, how customers make purchasing decisions. In the context of the choice-based RM framework, the behavioral models are in many cases extensions of discrete choice models whose properties have been extensively investigated. For the latter ones, small changes in model specification are expected to lead to results that do not significantly alter an assumed theoretical customer behavior. In contrast, given the novelty of the choice-based RM techniques, the same results may not apply to the proposed behavioral extensions. Specifically, due to how they are formulated and, subsequently, estimated these extensions may lose the interpretation stability of their original counterparts. To this end, an effort is made to understand how changes in the specification of the customer purchasing behavior model may impact the downstream processes such as the formulation of the preferred capacity control policy and the revenue performance of the overall choice-based RM system.

Objective 4: *Quantify the revenue gains attributable to the application of choice-based RM methodologies.*

According to the choice-based RM literature, it is expected that the use of the choice-based RM techniques will generate significant incremental revenues. To validate this claim, the revenue performance of the RM system currently in place at the hotel chain that sponsors this study is compared to that of a hypothetical choice-based replacement. A methodology that takes into account the conceptual differences between the two systems and simultaneously supports the direct comparison of their revenue performances has been devised. The corresponding competing capacity control policies are then determined by taking into account the observed demand for hotel products and its distribution during the booking horizon. These policies are finally used to evaluate the revenues that the hotel chain may gain/expect when similar arrival streams load onto competing systems.

1.3. Major Contributions

Within the choice-based RM literature there are two areas of practical contribution of this study. The first relates to situations that precede a potential implementation of a choice-based RM system. In these cases, interested organizations need to understand the data requirements that such an initiative commands and should plan for data collection systems and procedures that support its execution. The second refers to theoretical aspects that current choice-based RM methodologies appear to overlook. In particular, while they emphasize the importance of integrating the customer purchasing behavior into the standard optimization routines and suggest techniques to estimate the corresponding behavioral models, these methodologies employ simplistic model specifications that are not truly representative of how customers behave in the market place. In addition, the formulation and the estimation of these models lead to significant changes in the results reported when small adjustments to model specifications are

undertaken. In this context, we provide guidelines on how the model specification impacts all downstream RM processes including the overall revenue performance.

Compiling customer-centric data has taught us that collecting this type of data, post-processing it and checking for its validity is a daunting and time-consuming task. Despite the small number of hotel properties for which we have attempted to obtain customer-centric data, the amount of effort that was required to accomplish the task is illustrative of the difficulties that service providers would face when implementing a choice-based RM system (see Data Chapter for more details). The extensive use of advanced statistical techniques and outsourced services such as seeking the professional advice of experts employed in areas other than revenue management has revealed the fact that some firms may not be prepared from an institutional perspective to implement advanced RM systems. Therefore, it is our expectation that unless alternative data collection procedures are devised and successfully tested, some service operators and, in particular, those without highly trained staff will be reluctant to invest into choice-based RM systems simply because they do not have the capabilities to manage the expected data requirements. As a result, this study contributes to advancing our understanding of the choice-based RM methodologies by being the first study that identifies and quantifies the practical obstacles that service providers need to address prior to the implementation of advanced RM systems.

From a theoretical perspective, this study contributes to a better understanding of how the specification of customer purchasing behavior models impacts the recommended capacity control policies and the revenue performance of the corresponding choice-based RM systems. Our extensive testing on real data sets has shown that behavioral interpretations of the results returned by the Expectation - Maximization (EM) algorithm change from one model specification to another. In this context, the counter-intuitive estimates and the mixed result interpretations can be attributed to the formulation and estimation of the behavioral models. The instability of behavioral models' parameter

estimates impacts all downstream RM processes and generally leads to recommended capacity control policies that, while theoretically sound, are not completely representative of the markets they are supposed to serve. In such cases, the choice-based control policies tend to marginally outperform the traditional control policies used to manage a company's perishable products. The study of business environments in which customers purchase on price shows that when the customer purchasing behavior is understood the use of choice-based RM techniques leads to consistent and significant revenue benefits. In these environments, a theoretical alternative to the computationally intensive EM algorithm is proposed. This procedure provides for intuitive and simple means to account for the customer purchasing behavior while computing the recommended capacity control policies. By studying the behavioral models in the choice-based RM framework, we have learned that the RM field has not yet exploited the power of these models to its full potential. We believe that behavioral models are currently used as tools for providing a structured way to assigning and redistributing sets of purchasing probabilities required by the control optimizers rather than instruments that soundly describe the customer purchasing behavior. To accomplish a full integration of the customer behavior into the RM practices, it is our expectation that more advanced and/or alternative estimation and optimization techniques need to be devised.

1.4. Dissertation Structure

This dissertation consists of six chapters and one appendix. Chapter 2 reviews recent advances in the choice-based RM literature. This chapter also highlights areas in the existing methodologies which, due to limiting assumptions, may restrict their applicability. Lastly, this chapter emphasizes the importance of availability of customer-centric data for industry acceptance of choice-based RM practices. Chapter 3 describes the theoretical instruments used to build choice-based RM systems and compares their revenue performance against that of traditional systems. The relationships that govern

the interactions between the instruments that finally lead to the relative revenue performance assessment are formalized in a conceptual model described at the beginning of the chapter. Chapter 4 describes a data collection methodology devised to help the sponsor of this study collect customer-centric data. An in-depth look at the proposed data collection process provides new insights into practical issues that need to be addressed to successfully implement choice-based RM systems in the hospitality and other similar industries. The theoretical and empirical results of this study are presented in Chapter 5. The last chapter summarizes the major findings of this dissertation and provides directions for future research in the choice-based RM field. For completeness, a detailed description of some of the algorithms used to assess the relative revenue performance of competing capacity control policies is provided in Appendix A.

Chapter 2. BACKGROUND

When you talk about Revenue Management, people like the concept, but they have no idea how to effectuate the concept. There's a market out there for Revenue Management that's just unbelievable!

Herb Kelleher, co-founder, Chairman and former CEO
of Southwest Airlines (Cross [2])

2.1. Overview of Revenue Management Core Concepts

In the last two decades researchers and industry practitioners have proposed several definitions for Revenue Management. Amongst them, the most academic follows the principles of microeconomics and describes Revenue Management as being “the application of disciplined tactics that predict consumer behavior at the micro market level and optimize product availability and price to maximize revenue growth” (Cross [2]). While comprehensive in nature, this definition has not appealed to the revenue management community due to its involved technicality. To this end, in its most accepted formulation, Revenue Management is viewed as the process through which the right customers are offered the right products through the right distribution channel at the right time and the right price such that a firm’s revenues are maximized (American Airlines Inc. [24] as reported in Smith, et al. [25]).

To benefit from the application of the revenue management concepts, firms need to operate in a business environment that supports such an initiative. First, firms have to offer their customers a fixed, finite amount of products. A hotel operator, for example,

cannot sell more rooms than the room capacity of its establishment. Moreover, the hotel capacity is regarded fixed since changing it involves major capital investments that are not undertaken very often. Second, the products firms sell need to be perishable. This implies that products are of specific values until a precise moment in time. At the end of each business day, a hotel room, for instance, is of no value if the hotel operator fails to sell it to potential customers. Lastly, products sold need to appeal to customers with different product valuations. In this context, relative to a leisure customer who books her hotel stay well in advance of the arriving date, a business customer who requires an immediate room in a crowded hotel is expected to be willing to pay a premium to obtain the product of her choice.

Within a business environment that favors revenue management initiatives, firms are recommended to explore the principles of supply and demand economics to take advantage of the existing and/or future revenue opportunities. In particular, in a summary of the modern revenue management pillars extensively discussed in Cross [2], firms need to:

- Sell to micro markets, that is, firms have to have a mechanism in place to efficiently segment their customer base. The segmentation scheme is intended to group customers based on their purchasing behavior and product needs and, subsequently, help firms design products for each of the identified market segments. To avoid the spill-over of customers between products designed for different customer groups, firms need to enforce the market segmentation. Various strategies such as the fare restrictions adopted in the airline and hospitality industries (e.g., Saturday-night stay, advance purchase, non-refundable/exchange fees, etc.) can be employed to enforce the desired segmentation (for details, see Hanks, et al. [26], Kimes [27], Orkin [28] and Vinod [29]).
- Exploit product value cycles, that is, firms need to understand what customers in each segment value and how the value they associate with a given product and/or product

attribute changes in time (Matanovich, et al. [30]). The product value cycles acknowledge that a product consists of a set of attributes that meets a customer segment's needs. In addition, a good knowledge of the product value cycles helps firms better design their pricing strategies since customers' willingness to pay is theoretically known.

- Save products for valuable customers, that is, firms need to find who their most valuable customer segments are, how much product these customers request, and, equally important, when these customers request the product. The above information together with that associated with the product value cycles provides firms with the means to propose credible price points to stimulate demand from the price sensitive customer segments and, simultaneously, save enough capacity for the price insensitive customer segments.
- Focus on value-based pricing, that is, firms need to price their services and/or products based on the customers' willingness to pay rather than the cost they incur to provide the service and/or products. Value-based pricing requires firms to have a clear understanding of the dynamics of price elasticity by product and customer segment. In this context, Lewis and Shoemaker [31], Steed and Gu [32] and Varini, et al. [33] provide a comprehensive discussion of how the value-based pricing can be implemented in the hospitality industry.
- Focus on price when balancing supply and demand, that is, before implementing other, more radical solutions, firms need to adjust their prices to accommodate short-term demand fluctuations. In such cases, the prices that maximize firms' revenues by product and customer segment should be altered to account for the inherent variability in demand.
- Make decisions based on knowledge, that is, firms need to avoid the personal bias of human forecasters and should rely instead on mathematical algorithms to learn about the purchasing behavior of their customers. The downside of such an approach is that

the data required to support such complex initiatives is most of the time scarce, incomplete and not necessarily accurate.

- Continually re-evaluate revenue opportunities, that is, firms need to understand that the business environment in which they operate is dynamic. To take advantage of the changing revenue opportunities, firms need to be able to respond quickly to changes in the business environment which means that they have to approach the revenue management process from a system perspective.

2.2. Introduction to Choice-Based Revenue Management

The revenue management core concepts synthesized in Section 2.1 were articulated in the current form long after the existing revenue management (RM) systems were first implemented. Given that at their inception the impact of technological advances on their responsiveness could not be appropriately assessed and, therefore, could not be planned for, most of these legacy systems lack the comprehensiveness of the ideas presented above and continue to work with concepts that do not characterize anymore the today's changing market conditions. Recently, however, many researchers and industry practitioners have challenged some of the assumptions that guided the development of the existing RM systems and have opened the door for different approaches to solving the RM problems. The emerging concepts together with how they intend to overcome the limitations of the existing methodologies are all discussed in the remaining of this section.

Similar to what Cross [2] and other scholars (e.g., Lee [34], Talluri and van Ryzin [35]) have recommended, service providers such as airlines, hotels and car rental agencies tried from the beginning to appropriately segment their customer base and forecast the constrained and/or unconstrained demand associated with each customer segment. However, most RM implementations still operate based on the independent demand model (see Talluri and van Ryzin [35] for a detailed description of this model)

which assumes that “demand is associated with a product and is essentially independent of the market environment” (van Ryzin [15]). In short, the demand for a given product as computed through sophisticated forecasting procedures represents nothing more than a number whose value is unaffected by the competitive environment in which activities associated with it take place; that is, the availability of possible product substitutes - offered by the same firm or by one of its competitors - and their attributes are all ignored when estimating demand and, implicitly, when representing customer choices. In this context, the cause-effect contribution of environmental factors to the demand realization is viewed as pure noise by the current RM systems. Furthermore, in some applications of the independent demand model, the demand is assumed to arrive in a specific order, with the lower-priced product demand arriving early in the selling season and the higher-priced product demand showing up only after the former is realized. While analytically convenient and probably representative of how markets behaved decades ago, these assumptions fail to hold in the today’s fast changing business environment.

Facing threats as serious as lost revenues and revenue opportunities, the RM community has responded promptly and has started to investigate alternative approaches to dealing with RM problems. Building on the findings from previous studies (Belobaba [36], Brumelle, et al. [37], Pfeifer [38]) which acknowledged that customers’ purchasing decisions were indeed influenced by environmental factors (e.g., buy-up, buy-down, divert behaviors), the new theoretical developments have considered customers rather than products “as the fundamental unit of demand. After all, they do create demand in the first place. And understanding what influences their decision making is the key to building better models of demand” (van Ryzin [15]). This shift in thinking has generated high interest in modeling customer behavior in a more comprehensive RM framework. The most important advances in the resulting choice-based RM literature are summarized in Section 2.3 below. Given the mathematical complexity that most of the proposed approaches involve, Sections 2.4 and 2.5 discuss each of these individually and

investigate the appropriateness of the corresponding demand models. Since the airline industry has led most of these research efforts, the studies reviewed in the next sections come from this field.

2.3. Choice-Based Revenue Management: Summary of the Relevant Literature

As briefly mentioned in Section 2.2, within the traditional RM framework, the demand is typically associated with the products firms sell and not with the customers who purchase these products. As a consequence, when determining a preferred capacity control policy to manage product availability, the demand is forecast and, subsequently, analytically processed at product or product group levels. The resulting control policy specifies the products or product groups that are open for sale to every customer in each time period between two consecutive demand forecasts as well as the amount of capacity that is allocated and made available for sale to each of these product classes. In this context, as can be easily observed, determining the capacity control policy does not require firms to use any customer-related inputs.

Within the traditional RM framework, the first attempt to add a behavioral component to computing the capacity control policy is attributed to Belobaba [36], [39], [40]. Belobaba acknowledges that customers are willing to buy higher fare products, when discounted products are not available, and takes this explicitly into account when he determines the capacity control policy a firm needs to implement. In spite of promoting the behavioral integration, this approach does not fundamentally change the way the capacity control policy is determined; it does, however, alter the allocation of the available capacity across the product classes offered for sale. Through this improved allocation, firms are proved to gain significant incremental revenues.

More recently, in an attempt to capitalize even more on the potential benefits associated with the use of the choice-based RM systems, researchers have started to focus their efforts on alternative approaches to modeling customer behavior and appropriately

integrating it into the traditional RM practices. To this end, three distinct conceptual directions can be identified in the literature concerned with the choice-based revenue management. The first one attempts to incorporate customers' purchasing behavior into the RM framework and assess, at a macro level, the impact this integration has on the distribution of air travel demand across several competing airlines, itineraries and product classes. In contrast, the other directions focus on micro level RM activities, the point of interest being there to determine capacity control policies that maximize revenues at a particular airline.

As stated above, the first conceptual direction, which is primarily due to Belobaba (see Passenger Origin-Destination Simulator (PODS) described in Belobaba and Hopperstad [41], Belobaba and Wilson [42], Belobaba [43], and Carrier [44]), has looked at how different RM strategies that account for customers' purchasing behavior have influenced the performance of competing airlines. In this context, what makes the approach powerful is that the demand for product classes is the result of an exhaustive simulation of customers' choice that intends to replicate how real air markets function.

The second conceptual direction refers to researchers who have attempted to determine the structure of the preferred capacity control policy when customer purchasing behavior has been fully integrated into the RM system of a given airline. For example, continuing the traditional thinking outlined above but considering customers' preference for parallel flights, Zhang and Cooper [16] investigate how the inventory controls on these flights should simultaneously and profitably be set. In contrast, Bront, et al. [21], Gallego, et al. [17], Talluri and van Ryzin [14], van Ryzin and Liu [18] approach the same problem differently and consider that the demand is directly associated with customers requesting travel services. In this case, the resulting capacity control policy provides firms with the optimal sets of products that need to be shown to customers at each point in time prior to the departure date and for each amount of the remaining capacity such that firms' expected revenues are maximized. This means that

each customer could potentially face individualized consideration sets based on when she requests the service and what the inventory levels are when the request arrives.

Finally, the last conceptual direction identifies research which has considered that the firm employs a specific parametric capacity control policy whose parameters are determined such that the firms' performance is maximized. This is the approach proposed by van Ryzin and Vulcano [19], [20] and Vulcano, et al. [45] who based on the customers' preference for particular flights optimize through extensive simulations the parameters of the virtual nesting control policy they employ.

Most of the advances in the choice-based RM literature that were presented above involve complex mathematical concepts. In Sections 2.4 and 2.5 that come next, an attempt is made to briefly describe the general methodology that each of these approaches proposes. The methodological description is followed by an as in depth as possible discussion of the demand model considered and whenever opportune, the discussion is accompanied by comments on the effectiveness of the chosen demand model to appropriately represent customer behavior. Finally, since the approaches to incorporating customer behavior vary with the RM problem one wants to solve, both single-leg (Section 2.4) and network (Section 2.5) RM problems are considered. The single-leg RM problem investigates how the capacity of a single resource should be optimally/sub-optimally allocated to distinct fare product classes. For example, for a 100 seat flight from Atlanta, GA to Washington, DC the corresponding capacity control policy controls the availability of each of the fare products the airline sells (e.g., 15 seats are protected for the \$400 fare class, 45 seats are protected for the \$400 and \$300 fare classes, and the rest of the seats are made available to all \$400, \$300, and \$200 fare classes). The network RM problem, on the other hand, manages the allocation of the capacity of multiple resources across a network. To illustrate the point, if the flight from Atlanta, GA to Washington, DC connects other cities from the south (e.g., Miami, Houston, etc) with other cities from the north (e.g., Boston, New-York, etc.) then the

capacities on all flights (e.g., Miami - Atlanta, Atlanta - Washington, Washington - Boston, etc.) are simultaneously controlled by the corresponding capacity control policy.

2.4. Modeling Customer Behavior: Single-Leg Revenue Management Problems

For single-leg RM problems, Belobaba [36], [39] is credited with being the first one who added a behavioral flavor to the calculation of the capacity control policies (see also Belobaba and Weatherford [40] for a sound overview of these methods). His initial theoretical work was later validated by Bohutinsky [46] who, studying the buy-up phenomenon at Delta Air Lines, concluded that it in fact was present and was affecting disproportionately the various fare classes he investigated.

In his work, Belobaba [36], [39], [40] modified a widely used fixed protection level heuristic (i.e., expected marginal seat revenue heuristic EMSR-b) to account for the observed behavior of customers who, when discounted fare tickets were not available, purchased higher fare tickets and were therefore not lost for the airline company. From a RM system perspective, this correction meant that probabilities of various customers buying higher fare tickets when lower fare tickets were closed for sale were known entities. Belobaba and Weatherford [40] acknowledged that computing these probabilities was not a trivial task, and this observation made them cautiously state that “critical to the successful implementation and use of this new decision rule is the capability to provide accurate estimates of the sell-up [author’s note: buy-up] probabilities between price classes as inputs” (Belobaba and Weatherford [40], p. 362). Despite the difficulties associated with estimating the buy-up probabilities, based on the results of several simulated experiments, Belobaba and Weatherford concluded that “whether these probabilities are estimated statistically from historical booking data or simply provided as judgmental input by analysts knowledgeable about current market characteristics, it is clear that including customer diversion [author’s note: in this context, diversion, sell-up, and buy-up carry the same meaning] in the calculation of inventory

limits holds significant potential for incremental revenue gains [...]” (Belobaba and Weatherford [40], p. 362).

These conclusions may, however, be sensitive to the assumptions embedded in the model Belobaba and Weatherford proposed. In particular, they considered that (1) the demand for each fare product was independent of the demand for any other fare products; (2) the demand for each fare product during any booking period followed a parametric random distribution with known or estimable parameters; and (3) demand associated with any combination fare product - booking period was realized within that booking period. Furthermore, for each of the above fare - period combinations, a stationary Poisson process was assumed to appropriately describe customers’ arrivals (i.e., as pointed out by the authors, this approach was also used by Rothstein [47] and Williamson [48]). The mean arrival rates associated with these Poisson processes were further selected such that, across all booking periods in the booking horizon, “the commonly observed pattern of low-priced customers making their requests before customers willing to pay the highest prices” (Belobaba and Weatherford [40], p. 347) was reproduced.

While similar to the independent demand model with respect to some of the employed concepts (e.g., independence of demand across fare products), this model represented an advance over most previous research attempts in the sense that, within any given booking period, the arrivals of price sensitive and price insensitive customers were interspersed to account for the uncertainty in the order of arrivals. Moving away from the arbitrary rule of low-to-high order of arrivals constituted one of the first successful implementations of a dynamic single-resource model.

The modeling realism that came with this approach, however, required the introduction of additional constraints to an already heavily constrained problem. In particular, as described above, an assumption on the type of the arrival process was needed (e.g., Poisson process) to make this model tractable. To this, other

methodological issues that could question the direct applicability of this model are worth being mentioned. As pointed out by several other authors (e.g., Talluri and van Ryzin [35], [14]), the estimation of the buy-up probabilities between every two fare products in a multiple-fare-product situation was not practical due to the large number of such fare product combinations involved (e.g., a 10-fare-product situation requires the estimation of 45 buy-up probabilities). In addition, the probability of purchasing a higher fare product should vary in time, with customers being more willing to buy-up as the departure day approaches. In such a situation, the probability to buy-up to a higher fare product should depend on what other higher fare products are available at the time the intention to book is expressed and the characteristics of both the available products and the customer. Furthermore, even if a reasonable number of probabilities were to be computed, to date scholars in the field could not devise a methodology to effectively differentiate the customers that purchased a certain fare product from those who wanted to purchase the same fare product based on their original preference but, through diversion, ended up buying a different fare product. Thus, the estimates of the buy-up probabilities may be biased and reflect only partially the situation analyzed. Similarly, in spite of considerable efforts, researchers and industry practitioners have not yet provided the RM field with a sound theoretical answer to how these probabilities should be adjusted and/or recomputed when the competitive environment in which the firm performs changes (e.g., when a new competitor enters the market, the prices may drop and as a consequence the buy-up probabilities may need to be adjusted to better reflect the new competitive environment). It is our opinion that the mathematical complexity that all these efforts have involved represents one of the reasons why reasonable solutions could not be suggested.

As shown above, most of the work related to the single leg (static/dynamic) RM problem focused until late 1980s and early 1990s on developing easy-to-use heuristics (e.g., EMSR-b, EMSR-b with buy-up, etc.) to replace the more difficult exact

optimization models (e.g., dynamic programming, Monte Carlo integration, etc.). Talluri and van Ryzin [35] identified two reasons for why this situation perpetuated for so long. First, when the airline deregulation took place in mid 1970s the theory of the optimal controls was in its infant stage and not ready for a large-scale industrial deployment. As a consequence, all RM software packages that followed the deregulation incorporated these approximate solutions and made them available to the industry, which got devotedly attached to them. Second, heuristics are still popular especially among practitioners because there is no need for highly skilled personnel to build and maintain them. Furthermore, as practice consistently proved, these heuristics generate revenues that are most of the time similar to the optimal ones.

The popularity heuristics gained did not mean, however, that researchers stopped investigating alternative approaches. In particular, Curry [49], Wollmer [50], Brumelle and McGill [51], and Robinson [52] proposed optimal control policies for the static multiple-fare product RM problem. Lee and Hersh [53] and later Lautenbacher and Stidham [54] relaxed the fixed order of arrivals assumption and gave the optimal control policy for the dynamic multiple-fare product RM problem. Despite the sound theoretical concepts they were based on, these optimal policies followed the basic demand assumptions that Belobaba and others used in their work. As a consequence, the customer, the associated decision making process, and the competitive environment in which her decision takes place are all ignored during the optimization procedures. This counter-intuitive approach to the RM problem was supposedly corrected by Talluri and van Ryzin [14] who provided the exact solution to the single-leg model of RM under a general discrete choice model of consumer behavior.

Talluri and van Ryzin's approach to the single-leg problem of RM investigates how a firm can maximize its revenues by offering customers distinct sets of fare products based on the available capacity at the time the requests arrive and the time remaining until the product practically consumes. At a first glance, this procedure seems to be fairly

complex since, in order to determine the optimal set to display, one needs to search through all possible combinations of fare products at any given point in time (e.g., in theory, for a ten fare product situation, $2^{10}-1=1,023$ product combinations should be investigated). Despite this apparent complexity, Talluri and van Ryzin show that only a small sample of all fare product combinations determines the optimal control policy. Specifically, they prove that only the efficient subsets, that is, the sets that “provide the most favorable trade-off between total probability of purchase and expected revenue” (Talluri and van Ryzin [14], p. 16) are used to find the optimal control policy.

This policy, however, needs to be determined and implemented in the RM system prior to the actual arrivals (i.e., prior to the start of the sale process) and, therefore, to facilitate its deployment, certain assumptions about the customer purchasing behavior need to be explicitly made. In particular, for each arrival, a set of probabilities associated with the customer buying each of the fare products offered and, for that matter, none of them needs to be specified into the optimization procedure that computes the optimal order of displaying the efficient subsets. Based on transaction level data enriched with time of booking information, both simulated, Talluri and van Ryzin provide a general methodology to estimate a multinomial logit model (MNL) of customer choice and, subsequently, they use this model as a consistent way to determine the sets of required probabilities. The independence from irrelevant alternatives property of the MNL models together with the particular way their model was specified (i.e., no alternative specific constants were estimated, price was the only explanatory variable considered) assure that parameter estimates of this model can be used to determine the probability sets even if the cardinal of these sets changes in time (i.e., the procedure is applicable even if the efficient sets shown to customers change their product composition as the departure date approaches and the available capacity decreases). Moreover, for the MNL choice models, Talluri and van Ryzin show that the “optimal control policy is a nested allocation policy where the nesting is by fare order” (Talluri and van Ryzin [14], Proposition 6, p.

25). In short, this means that the currently offered efficient subset is ordered by fare, is complete, and is contained in all efficient subsets previously displayed.

An illustration of this concept is provided in Figure 1 below. There, at the beginning of the booking horizon, discretized in T elemental time periods, all five fare products are available to all potential customers. The efficient subset S_S which consists of all products is therefore displayed at the time $t = T$. As the time passes and the available capacity decreases, only a few of the products will be made available for sale at an intermediate time t (e.g., efficient subsets S_{s+1} , S_s , and S_{s-1}). It is important to note that all efficient subsets S_s , $s = S, \dots, 0$, are ordered by fare, that is, the first element in subset S_s is the most expensive product, the second element in the subset is the second most expensive product, etc. In addition, all efficient subsets S_s are complete in the sense that, if the cardinal of subset S_s is k , then subset S_s consists of the first k^{th} most expensive products. Finally, efficient subsets S_s meet the containing requirement

$$S_0 \subseteq \dots \subseteq S_s \subseteq \dots \subseteq S_S.$$

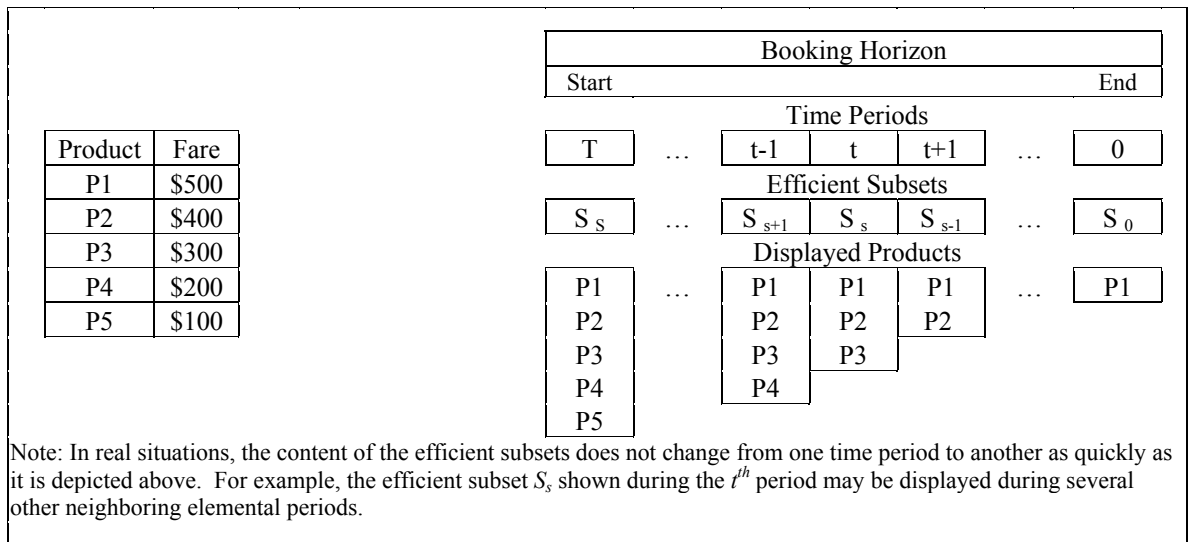


Figure 1 Optimal Control Policy - Nesting by Fare Order

The methodology presented in Talluri and van Ryzin [14] represents the first attempt intended to soundly integrate consumer purchasing behavior into the RM framework. The expectation-maximization method that it proposes to estimate the MNL

model is very powerful in that it takes into account the random cause of a selling period with no observed purchases when it attempts to compute the model parameter estimates. The proposed approach to quantitatively describing the behavioral model is needed since very often service providers cannot discriminate in their purchase transaction level data between periods of time with no-arrivals and those with arrivals and no-purchases. Given this limitation, the traditional maximum likelihood estimation method which would normally be employed cannot be used to compute the desired parameter estimates.

Similarly, the integration of the MNL model with the optimization routine that determines the optimal capacity control policy is also innovative. As such, even if the MNL model is applied repeatedly to choice situations that involve a varying number of alternatives, its initial specification does not require the analyst to refine or adjust in any way the parameter estimates once they are determined. While it offers the methodology consistency and generality, such an approach may, however, lead to questionable results. Omitting the alternative specific constants from the model specification allows the analyst to easily determine the outcome of any choice situation but, in the same time, significantly reduces the explanatory power of the model. In addition, most likely because of the product-oriented character of the (simulated) data they worked with, Talluri and van Ryzin consider that the consumer choice is influenced by the product attributes only (i.e., product price); due to their omission, customers' socio-demographic characteristics and the competitive environment in which purchasing decisions take place are not considered to contribute in any way to the overall decision making process. Finally, customers are most of the time heterogeneous in their behavior (e.g., price sensitivity varies across customers) and, therefore, a simple MNL model may not appropriately describe this phenomenon. As a consequence, more advanced models that allow random taste variation across individuals may be more appropriate to handle this situation.

2.5. Modeling Customer Behavior: Network Revenue Management Problems

The single-leg RM problems that incorporate customers' purchasing behavior and that were discussed so far are not difficult to operationalize. In contrast, modeling customer behavior on networks leads to more complex problems and, as a consequence, the number of simplifying assumptions employed when trying to answer those increases. Given that these assumptions are most of the time briefly described in the literature, in the next paragraphs, an attempt is made to identify those aspects of the reported network RM models that may generate problems and describe their likely behavioral implications.

As in the case of the single-leg RM problem, it is believed that the first who considered customer choice behavior in network problems was Belobaba (Belobaba and Hopperstad [41]). Since the Passenger Origin-Destination Simulator (PODS), developed by Boeing and enhanced with RM simulation capabilities at the MIT Flight Transportation Laboratory, is used repeatedly in Belobaba's work (for specific details see Belobaba and Hopperstad [41], Belobaba and Wilson [42], Belobaba [43], and Carrier [44]), an attempt to describe its Revenue Management and Passenger Choice & Booking modules is made here. However, before moving to this task, it may be worth mentioning several things that will help one better understand Belobaba's research methodology.

First, for the purpose of PODS, customers are grouped in two distinct classes: "business travelers and leisure travelers, each with different sensitivities to differences in price and attributes of fare products, as well as to the departure times and schedule frequencies of competing carriers" (Belobaba and Wilson [42], p. 4). In addition, each customer is assumed to be a utility maximizer in the sense that, when faced with multiple options, the customer buys the itinerary-fare product on the flight that maximizes her utility of travel, which includes both the out-of-pocket cost and the inconvenience of travel costs (i.e., the inconvenience associated with product restrictions, the departure from the desired flight schedule, the quality of the itinerary and the seat unavailability at

the preferred airline are all converted into monetary units). Furthermore, Belobaba considers that an arriving customer always selects a product from a set of available itinerary-fare products whose prices do not vary in time^{2, 3}.

With these basic concepts detailed, we can move now to presenting the revenue management module employed in PODS studies. Based on simulated historical booking data, the RM module forecasts demand for each itinerary-fare product at the departure date level and, subsequently, computes a set of booking limits at the beginning of each booking horizon. The booking limits control the amount of available capacity that can be sold to any itinerary-fare product category and remain in place until a new demand forecast is employed. Then, based on the amount of capacity already sold and the expected demand until the departure date, the set of booking limits is revised and used until the next demand forecast takes place. This process continues with increasing frequency until the considered flight departs. For each of the time intervals between two successive demand forecasts or timeframes if we were to use Belobaba's terminology, the passenger choice and booking module is employed to simulate the arrival of customers and, implicitly, their purchasing behavior⁴. To accomplish this, the proportion of the total forecast demand that is expected to come in the current timeframe, by passenger

² The constant price over time assumption was true at the time these studies were undertaken. However, in the today's competitive environment, this assumption does not hold anymore with airlines, for example, changing their ticket prices several times during a business time interval.

³ In PODS studies, each customer has to choose among, at most, 25 different itinerary-fare products (i.e., for each origin-destination market, there are 2 airlines, 3 itineraries per airline, 4 fare products per itinerary which result in 24 distinct itinerary-fare products to which the no-go alternative is added). This number varies, however, across customers based on *the RM controls* (e.g., if at the time of booking, a given itinerary-fare product class is closed, then the corresponding product will not appear in customer's choice set), *the advanced purchase restrictions* (e.g., if the customer books 5 days in advance of the departure date then all the itinerary-fare products with more restrictive advanced purchase requirements will not show in customer's choice set), and *the willingness to pay* (i.e., itinerary-fare products that exceed the simulated customer's willingness to pay do not enter her choice set). Given that customers are assumed to be utility maximizers, they will *always* select the itinerary-fare product that will provide them with the highest utility; in this framework, as imposed by the authors, the no-go alternative is selected if and only if it is the only alternative available (Carrier [44], p.55).

⁴ The booking horizon starts 9 weeks before each of the simulated departure dates and consists of 16 time frames. Initially, the time frames last a week but as the time passes and the departure date approaches they shrink to better describe a more intense booking activity associated with the end of the selling period (Carrier [44]).

type, is first obtained based on an “input ‘booking curve’ that determines the probability that, for example, a leisure passenger will book in each timeframe of the booking process” (Belobaba and Wilson [42], p. 5). The timeframe demand is then scrambled, to account for a random passenger arrival order, and the itinerary-fare product choice associated with each arrival is next simulated based on several characteristics. Specifically, customer choice is assumed to be determined by attributes such as: (1) customer’s willingness to pay, (2) perceived monetary costs associated with various fare product restrictions, and (3) perceived monetary costs associated with other, less attractive flight options. These attributes are randomly drawn from distributions that characterize each passenger type and, when combined, they determine the itinerary-fare product that the customer selects. Once the customer makes her decision to buy, the remaining capacity is updated to reflect the sale event and the process continues similarly for all expected timeframe arrivals. The entire process described above is schematically depicted in Figure 2 below.

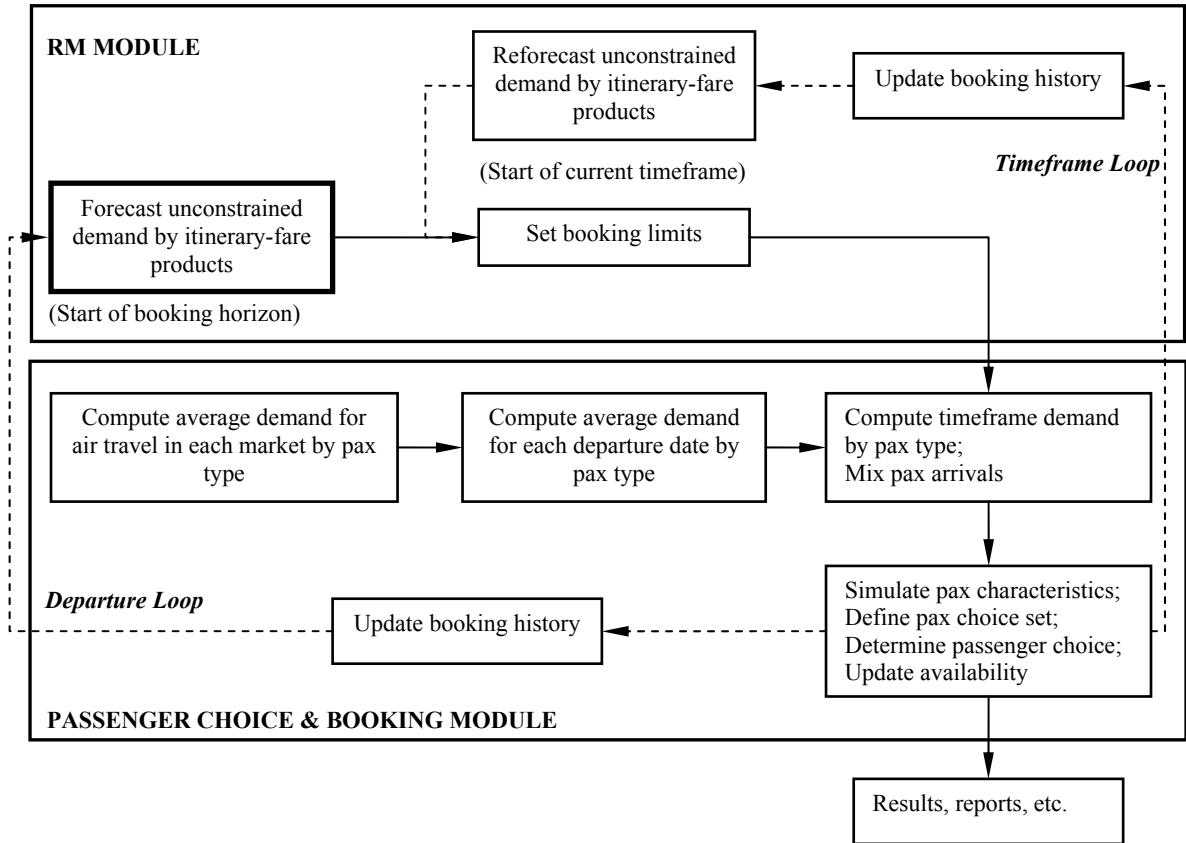


Figure 2 PODS System Flow [Original in Hopperstad [55], p. 5; Adapted from Belobaba and Wilson [42], p. 5]

For the PODS system shown in Figure 2, a few comments are worth being made. First, the average demand for air travel in each market is computed through a gravity model that takes into account the attractiveness of each origin-destination pair in that market. The predictive capabilities of the gravity model are, however, “unclear, especially in light of its explicit lack of behavioral assumptions” (Meyer and Miller [56], p. 281) and, therefore, such a model should be applied with care. Second, the parameters of the distributions from which the random characteristics are drawn need to be specified in advanced of the execution of the simulation program. This means that the analyst knows quite well the amount of out-of-pocket money that customers are willing to pay for their tickets and the monetary costs that they associate with ticket restrictions and the inconvenience of flight alternatives other than the desired ones. While not impossible to

construct, such a procedure is sensitive to the interdependencies that exist among the considered attributes (e.g., leisure travelers who book the lowest fare products tend to have low monetary costs associated with the advance purchase and connecting flights restrictions) and needs to be carefully approached.

The approach to modeling customer behavior in network RM problems proposed by Belobaba had a significant impact on the airline industry. Its success may be partly attributed to the way airlines were structured and functioning prior to the tragedy of 9/11/2001. The fact that business customers could have shown up at the boarding gate minutes prior to the departure and requested an expensive seat for that flight supported the applicability of Belobaba's assumptions. However, in the context of the operating procedures that airlines currently have to conform with, the clear distinction between business and leisure customers deteriorated and segmenting customers in two groups only became an assumption that most airline operators were forced to disregard or readapt.

Within the new stream of research, the work of Zhang and Cooper [16] investigates how the inventory controls on different flights (i.e., the booking limits associated with each fare product) should simultaneously and appropriately be set when a given airline flies multiple single-leg parallel flights within a short period of time⁵. Here, for a set of n available flights, m possible fare product classes, each characterized by its fare $f_j, j=1, \dots, m$, are assumed to be sold. In addition, requests for a fare product class j are grouped together and expected to arrive only in time period j . Given that in this period, customers face the same fare products irrespective of the flight they choose, the problem of interest is to find the control policy that satisfyingly assists the distribution of customers among the available flights (i.e., customers are assumed to (1) accept currently

⁵ *Parallel flights* refer here to flights that are flown between a particular origin and destination in a short period of time. For example, Delta Air Lines (www.delta.com, Access Date: 10/01/2006) flies four flights from Atlanta - Hartsfield-Jackson, GA to Washington - Reagan National, DC Monday mornings between 6am and 10am.

offered fare product and (2) not have a strong preference for only one of the available flights).

To determine the preferred control policy, Zhang and Cooper formulate their problem as a Markov decision process and consider that customers select their favorite flight according to a *preference mapping* choice model. The details of this model are succinctly given in the followings. During a given time period, the requests for service are assumed to arrive randomly. In addition, each customer is considered to be characterized, prior to her arrival, by a preference mapping that is used to make the decisions of whether or not, and what to buy (e.g., for a mapping function of the form $\theta: N_o \rightarrow M$, flight i is said to be preferred to flight j if $\theta(i) < \theta(j)$)⁶. These decisions are inevitably influenced by the seat availability associated with each flight i . Faced with making a purchase decision, a customer will buy a seat in the flight most preferred given that the seat is available. If it is not, the customer will buy the seat in the second most preferred flight if a seat is available for her, and the process repeats until all flights are investigated or the *no buy* option is incurred. For the hypothetical example presented in Footnote 6, a possible choice scenario is shown in Figure 3 below. Here, given that flight 3 is full, the customer cannot buy a seat on her most preferred flight. However, because seats are still available on flight 2, she purchases a ticket for this flight. Had flight 2 been full as well, the customer would have decided to not take the trip. Similar to the other simulation studies discussed so far, the preference mapping for customers arriving in time period j is controlled by the analyst who imposes it based on some predetermined criteria (see Zhang and Cooper [16], Section 9, p. 427-430 for specific examples).

⁶ To illustrate the preference mapping concepts let's consider the following example: a customer has the option to select among three competing flights and before she investigates the seat availability on each of the flights she decides that she is going to buy either on flight 3 or flight 2, necessarily in this order. If there are no seats available in any of these two flights, she decides not to buy. For this customer, the mapping function $\theta: N_o = \{0, 1, 2, 3\} \rightarrow M = \{1, 2, 3, 4\}$ takes on the following values: $\theta(3)=1$ (customer's first option is flight 3), $\theta(2)=2$ (customer's second option is flight 2), $\theta(0)=3$ (if there are no seats available on flights 3 and 2, do not buy; 0 = no buy), and $\theta(1)=4$.

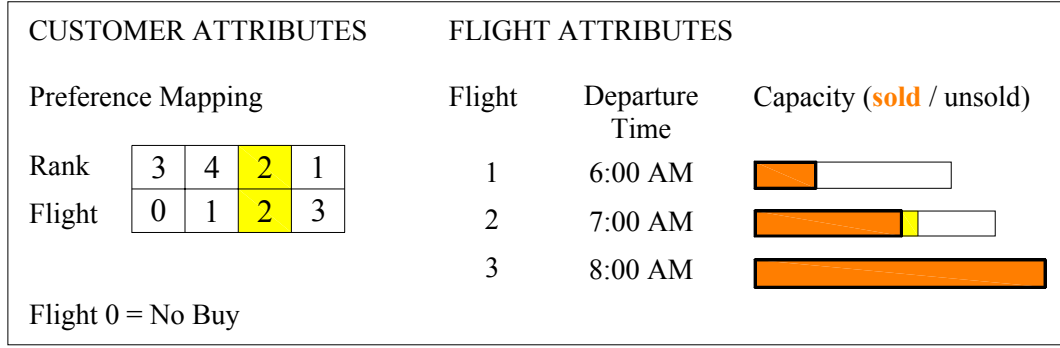


Figure 3 Customer Flight Selection - Preference Mapping & Seat Availability

The model of Zhang and Cooper, as complex as it appears, does not meet the realism required for a large industrial implementation. Specifically, it considers that the entire demand for a fare product class j across all available flights arrives randomly in time period j only⁷. This assumption does not take into account the uncertainty in the order of arrivals and, therefore, imposes inflexible behavioral limitations. In addition, the proposed methodology does not allow customers to buy up on their preferred flight if their assigned class is closed and seats are available in higher fare classes. While this may be applicable to certain leisure travelers who do not have to adhere to rigid schedules, this does not apply to business customers who are, most of the time, willing to pay for extra convenience (i.e., fly the preferred flight no matter what the fare is) rather than take a different flight. As a final comment, in the real world, airlines may use strategies, complementary to seat availability controls, to properly distribute demand among competing parallel flights. For example, if an 8:00 am flight is considered to attract more business travelers than leisure travelers, the airline may discount the tickets for the 6:00 am competing flight to stimulate the leisure demand for that flight and free as much capacity as possible on the 8:00 am flight for customers who are inflexible with respect to their flight preferences.

⁷ The assumption that the demand for fare product class j arrives during time period j only suggests that there exist distinct customer segments that are adequately separated by the class restrictions airlines use. Given the competitive environment in which airlines compete, this assumption is regarded nowadays with real skepticism (van Ryzin and Liu [18]).

The solutions Zhang and Cooper propose in their work refer to the special case of network parallel flights. van Ryzin and Liu [18] acknowledge this shortcoming and recommend a more general approach to solving the network RM problem. In particular, van Ryzin and Liu formulate firm's decision problem as a dynamic program (DP) whose solution identifies, at each time t prior to the consumption of the product and for each possible state of the network, the optimal set of products S to offer for sale. They recognize, however, that the DP problem is not solvable for networks of reasonable size and, based on the earlier work of Gallego, et al. [17], they review and advocate the use of the choice-based deterministic linear programming (CDLP) approximation.

The CDLP method assumes that demand and capacity are continuous (as opposed to discrete) entities and that, for each product set S that may be offered to customers, its expected revenue $R(S)$ and consumption rate $Q(S)$ are deterministically known. With these assumptions in place, the original decision problem changes and becomes that of finding $t(S)$, the total number of periods in which product set S is offered. The computational efficiency associated with this transformation is, however, accompanied by a certain amount of ambiguity, in the sense that, the optimal solution of the CDLP problem gives the total time $t(S)$ but not the order in which sets S should be presented to customers (for more details see van Ryzin and Liu [18]). In spite of this inconclusiveness, it is worth mentioning that in the optimal solution only $(m+1)$ product sets S of the (2^n-1) possible non-zero ones are present, which indeed represents a major computational achievement⁸. To compute the CDLP efficiently, van Ryzin and Liu use the column generation technique (for specific details see Winston and Venkataramanan [57], p. 570) , which is briefly described next.

⁸ In this context, m and n refer to the number of legs in the network and the number of products the firm sells, respectively. Given that the n products can be grouped in maximum (2^n-1) distinct non-zero sets, (2^n-1) variables $t(S)$ will enter the CDLP maximization problem. However, due to the $(m+1)$ constraints that condition the values of $t(S)$'s, only $(m+1)$ such sets S will appear in the optimal solution.

Conceptually, only a small number of product sets S is first considered to enter the CDLP revenue maximization program. This reduced problem is then solved based on the information associated with sets S and the resulting dual solution is used to investigate whether the introduction of other sets S will increase the objective function of this initial CDLP program. If an increase in the objective function is guaranteed then the corresponding product set S_k is added and the CDLP program is solved again. If, on the other hand, such an increase is not confirmed then current sets S give the optimal solution and the iterative process stops. Using this approach, the set of optimal product sets S expands as the algorithm advances, “the hope being that only a modest number of columns [A/N, sets] needs to be generated before optimality is reached” (van Ryzin and Liu [18], p. 15). As previously mentioned, the optimal CDLP solution fails to provide one with the correct order in which optimal sets S should be displayed. To overcome this issue, van Ryzin and Liu propose a sophisticated approach that uses the optimal dual CDLP solution to decompose the network RM problem into a sequence of single-leg RM problems which are much easier to solve (see Talluri and van Ryzin [14] for details on how to solve the single-leg RM problem). In particular, for each leg i of the network, through the approximation of the network value function $V_t(x)$, the one-dimensional leg value functions $V_t^i(x_i)$ are independently determined⁹. This means that, at each time t during the booking horizon, the opportunity cost of a unit of capacity on leg i is known and is statically expressed in terms of the value of capacity elsewhere in the network. The leg value functions $V_t^i(x_i)$ are then combined to recreate a transformed network value function, which is used to dynamically determine the sequence of optimal sets S to be shown to customers.

⁹ A generic value function $V_t(x)$ gives the optimal expected revenue in the system at time t as a function of the remaining capacity x . The value function is used to compute the opportunity cost (or displacement cost) of capacity which is then compared with the revenue associated with a request and used to accept or deny the request.

As described above, the CDLP methodology involves techniques that are computationally expensive. Even if not explicitly incorporated in our previous discussion, van Ryzin and Liu [18] consider that the way customers' choice is modeled has a significant impact on the attractiveness, if not appropriateness, of their approach. In particular, they recognize that “the complexity of solving the [...] offer set selection problem depends on the choice model” (van Ryzin and Liu [18], p. 17) one employs. As an example of an efficient way to handle the entire CDLP problem, van Ryzin and Liu discuss the particular case when the probabilities associated with customers purchasing products from offered sets S are given by the MNL with disjoint consideration sets model. In this framework, customers are grouped in L distinct segments, each of which is identified by a unique consideration set S_l . These consideration sets are assumed disjoint, that is, for any two indices m and n , the intersection $S_m \cap S_n = \emptyset$ if $m \neq n$. In addition, customer arrival rate λ is assumed to additively distribute among the L customer segments or $\lambda = \sum_{l=1}^L \lambda_l$. For this specific choice model, van Ryzin and Liu theoretically show that solving the CDLP problem by column generation and, subsequently, determining the optimal order of display are operations that can be efficiently solved. Furthermore, they report encouraging results obtained from several simulated experiments (e.g., for a parallel flight experiment, the application of proposed methodology leads to revenue gains on the order of 1-5%).

In spite of this apparent success, a few comments are worth being mentioned. First, the arrival rates λ_l in all simulated experiments are completely specified by the analyst and not estimated from either real or synthetic datasets. This means that the analyst can perfectly segment her customer base, an assumption which is rarely true in real applications. Second, for each segment l , the preference weight associated with each product in the consideration set S_l (i.e., the antilogarithm of the product observed utility) is also arbitrarily specified by the analyst. While a knowledgeable analyst may know

quite well how customers behave in a given market, she cannot successfully replace the rigor of an estimated choice model. Since the results may be sensitive to how the preference weights were defined, the reported revenue gains should therefore be interpreted with caution. Lastly, it should be mentioned that given the proposed segmentation scheme, only L choice sets out of the $2^n - 1$ possible ones, are taken into consideration when searching for the optimal solution (to some extent, this reminds of the efficient subsets described in Talluri and van Ryzin [14]). As a consequence, while appropriate for networks with customers that can be appropriately segmented, this approach may not perform as expected in markets with heterogeneous customers.

The work of Zhang and Cooper [16], Gallego, et al. [17], and van Ryzin and Liu [18] “are similar in that each tries to determine (or approximate) the structure of a choice-based network capacity control policy. That is, they do not assume a policy *a priori*, but rather the policy structure is an output of their analysis. To achieve this, however, requires making simplifying assumptions or approximations, e.g., that demand is deterministic as in the LP [author’s note: linear programming] model analyzed by Gallego, et al. [17] and van Ryzin and Liu [18], or that the network consists exclusively of parallel flights with customers choosing among alternatives within the same fare class as in the work of Zhang and Cooper [16]” (van Ryzin and Vulcano [19], p. 3). To lessen the implications of these simplifying assumptions, van Ryzin and Vulcano [19], [20] propose a different approach to solving the network RM problem. In particular, they consider that the firm employs a specific *parametric* capacity control policy whose parameters are optimized through extensive simulations. In addition, the optimization routine connects with an independent demand module that seems to be able to handle any “choice behavior and any demand model one can simulate” (van Ryzin and Vulcano [19], p. 3).

The demand module provides the optimization routine with multiple *streams of demand* or *sample paths*, each of which is identified by the number of arriving customers,

the arrival order, the quantity of product requested by each customer and, most notably, customers' product preferences. To illustrate, each stream of demand can be viewed as a sequence of the form $w = \{(L_1, Q_1), (L_2, Q_2), \dots, (L_t, Q_t), \dots, (L_T, Q_T)\}$ where T specifies the number of customers, the index $t \in [1, T]$ gives the order of arrival, Q_t , a continuous measure, specifies the quantity of product that each customer requires, and $L_t = [L_{t1}, \dots, L_{tk}, \dots, L_{tN}]$ associates for each customer t a discrete preference rank k to each product j (i.e., $L_{tk} = j$)¹⁰. As in most already discussed applications, the difficulty here is to appropriately specify the ranking vector L_t . van Ryzin and Vulcano [19], [20] recognize this difficulty and, for all their numerical examples, they assume that customers can be categorized in several customer segments, with each segment having its own predetermined preference order for the products. As before, the number of customer segments, the preference and the arrival order are all arbitrarily specified by the analyst in a way that fits best the requirements/conditions of the particular problem that she studies¹¹. In a more general context, this situation is best described by van Ryzin and Vulcano who conclude their demand model section with “in essence all we require is an **‘oracle’** which can generate sample paths w drawn from some (perhaps implicitly defined) distribution. For example, the sequence and preference could be obtained via a detailed simulation of each individual customer's decision processes as in the PODS simulations of Belobaba and Hopperstad [41]” (van Ryzin and Vulcano [19], p. 7).

¹⁰ If we revisit the example of Footnote 9, for the customer for whom we detailed the mapping function θ , we can write the ranking vector L_t as $L_t = [L_{t1}, L_{t2}, L_{t3}, L_{t4}] = [3, 2, 0, 1]$. As before, this means that customer's 1st preferred choice is to book a seat on the 3rd flight, her 2nd preferred choice is to buy a seat on the 2nd flight, and her 3rd preferred choice is to purchase on none of the flights. For completeness, we should mention that L_t appears to be the inverse of the mapping function θ , or $L_t = \theta^{-1}$.

¹¹ A *buy-up* problem, for example, involves the classical scenario of two fare classes (i.e., low and high) and three types of customers. Customers type 1 are assumed are only willing to buy the low fare, those type 2 are considered are only willing to buy the high fare and those type 3 are assumed to prefer the low fare but buy the high fare if that is the only one available. In addition, customers type 1 and 3 are assumed to arrive first, at random, followed by customers type 2.

2.6. Summary

The review of the choice-based RM literature shows that modeling customer purchasing behavior in RM applications has passed the exploration stage. Slowly but surely, the limiting assumptions that accompanied the initial RM developments are replaced with other, less restrictive and more reliable ones.

To this end, as the field matures, it can be noted that researchers try to move away from the classical product-level models of demand and focus instead on behavioral such models. The shift in thinking of customers as the elemental unit of demand is, however, not simple. Indeed, all methodological approaches that explicitly consider customers and not products as demand generators rely on simplifying, analyst induced assumptions that limit to some extent their applicability. Talluri and van Ryzin [14], for example, use a simple MNL model to represent the decision process through which customers go when faced with a multiple product choice situation. Similarly, Zhang and Cooper [16] assume that customers who require seats on routes served by competing parallel flights stay with their pre-assigned fare classes when they select their choice of flight. While Gallego, et al. [17] work with demand streams of similar customers, van Ryzin and Liu [18] acknowledge that customers differ with respect to various attributes and segment them in classes with disjoint consideration sets. In the same context, Bront, et al. [21] construct the (sub)optimal product display policy based on only this reduced number of consideration sets, but allow for overlapping customer segments. Finally, each customer segment used in van Ryzin and Vulcano [19], [20] is associated with a predetermined preference order for products which is specified *a priori* by the analyst. The methods these authors employ in their work together with some of the limitations of their approaches are all summarized in Table 1 below.

To the methodological limitations that inherently accompany the development of all cutting-edge concepts, obstacles of a different nature contribute as well to the slow

acceptance of the choice-based RM techniques. In particular, the quantity and the quality of customer-centric data currently available to the research community do not satisfy the data requirements of the technical approaches it proposes. As a result, researchers in the RM field resort to simulated data to prove the revenue superiority of their techniques. The promising incremental revenue gains reported as outcomes of these studies do not, however, answer the viability concerns of an entire community of skeptical industry practitioners. Given that the successful implementation of a large scale, choice-based RM system has not yet been reported, the theoretical benefits of such an approach cannot be confirmed, and, as a consequence, they are hard to sell. “But an increasing amount of these data [i.e., customer-centric data] is becoming available, the quality is improving, and the technology to manage it is constantly advancing. So what appears to be an insurmountable obstacle today could well become practically feasible in the near future” (van Ryzin [15], p. 208).

Table 1 Summary of Relevant Choice-Based RM Literature

Authors	Method / Approach	Characteristics of the Demand Model
Single-Leg RM Problems		
Belobaba [36], [39], Belobaba and Weatherford [40]	EMSR-b with buy-up, Heuristic / Simulated Experiments	Imposed arrival types to maintain model's analytical tractability (e.g., Poisson processes). Limited guidance on how to compute the buy-up probabilities in multi-product and/or dynamic competitive environments.
Talluri and van Ryzin [14]	Dynamic Programming, Exact Solution / Simulated Experiments	Simple specification of the customer purchasing behavior model. Limited testing of the performance of the EM algorithm.
Network RM Problems		
Belobaba and Hopperstad [41], Belobaba and Wilson [42], Belobaba [43]	PODS, Heuristic / Simulated Experiments	Conversion of product demand to passenger type demand may be improved. Conversion of customer and product characteristics to monetary costs may be refined.
Zhang and Cooper [16]	Markov Decision Process, Exact Solution and Heuristics / Simulated Experiments	Demand for a product associated with a specific time period. Redistribution of demand among product classes may be improved.
Gallego, et al. [17]	Deterministic Linear Programming, Exact Solution and Heuristics / Simulated Experiments	Constant arrival rates over the entire booking horizon. Product prices considered exogenous. Demand model does not support customer segmentation.
van Ryzin and Liu [18]	Dynamic Programming, Exact Solution and Heuristics / Simulated Experiments	Customers segmented in classes with disjoint consideration sets. Pre-specified arrival rates (vs. estimated arrival rates). Pre-specified preference weights (vs. estimated preference weights). Solution investigates a limited number of product sets.
van Ryzin and Vulcano [19], [20]	Stochastic Approximation Algorithm, Exact Solution and Heuristics / Simulated Experiments	Customers segmented in several customer classes each with its predetermined preference order for products. The number of customer segments, the arrival order, and the product preference specified by the analyst.

Chapter 3. CONCEPTUAL MODEL AND METHODOLOGY

Man tries to make for himself in the fashion that suits him best a simplified and intelligible picture of the world; he then tries to some extent to substitute this cosmos of his for the world of experience, and thus to overcome it.

Albert Einstein, German-born theoretical physicist

3.1. Introduction

This chapter details the conceptual model that guides the development of a choice-based RM system and assists in quantifying its revenue performance. Since such a RM system accounts for the customer purchasing behavior when recommending a preferred capacity control policy, it necessarily builds on knowledge learned from processing customer-centric data. The minimum data requirements that are needed to promote a choice-based RM implementation together with a methodology devised to support the collection of such data are briefly introduced in Section 3.2 that comes next. The techniques that allow the estimation of a rational customer purchasing behavior model and, later, the recommendation of an optimal product display policy are also presented. The introductory concepts of Section 3.2 are subsequently elaborated in the next sections as follows: Section 3.3 expands on the concepts associated with the introduction of the conceptual model; Section 3.4 complements the classical discrete choice theory and provides an efficient way to deal with the incompleteness of the purchase transaction level data most service operators work with; Section 3.5 introduces the EMSR-b heuristic which is widely used in several industries to control the availability of perishable

inventories and describes the details of computing an optimal product display policy based on the customer purchasing behavior previously modeled; and, finally, Section 3.6 proposes a methodology that evaluates and compares the revenue performance of the competing capacity allocation policies detailed in Section 3.5.

3.2. Conceptual Model - Abstract Representation

The choice-based RM literature reviewed in Chapter 2 emphasizes the potential benefits associated with the integration of the customer purchasing behavior into the RM practices. The conceptual model used to study the legitimacy of these claims is depicted in Figure 4 and involves processes grouped into three distinct categories. At the first level, the customer-centric data and the algorithms used to process it and recommend behavioral control policies form the foundation for the implementation of choice-based RM systems. At this first level a methodology that allows a hospitality service provider to collect customer-centric data is proposed and further detailed in Chapter 4. This methodology, slightly modified to account for industry particularities, can be used by other interested parties to meet the data requirements of a choice-based RM implementation. The algorithms that facilitate the conversion of the raw customer data into actionable capacity decisions are generally not made publicly available. As a result, customized scripts are written to assist with the computation of behavioral-based capacity control policies. These scripts account for the incompleteness of the available customer-centric data and are, therefore, able to accurately describe the customer behavior. At the next level, an attempt is made to assess the relative revenue performance of choice-based RM systems. In this context, procedures that permit the direct comparison of the performance of competing control policies are formulated and incorporated in automated decision support tools. The last stage of the conceptual model includes implications and limitations of the methodology proposed in this study. We also emphasize here several issues related to the generalizability of the findings to other industries and contexts.

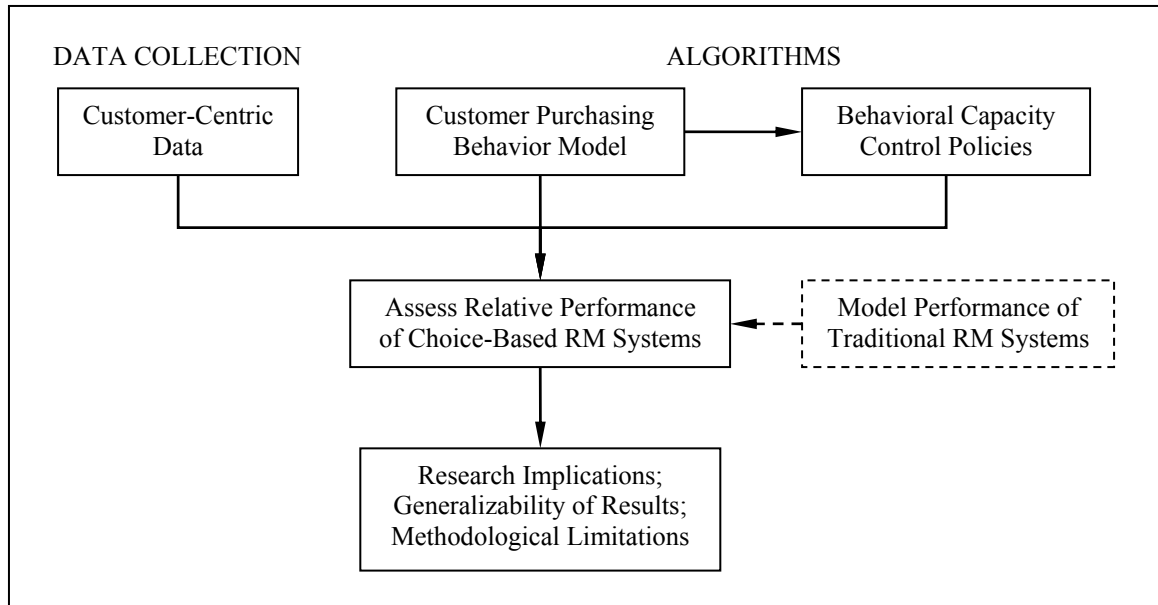


Figure 4 Conceptual Model

3.3. Conceptual Model - Details and Supporting Arguments

This section builds on the concepts introduced above and details the steps that need to be undertaken to successfully build a choice-based RM system and accurately assess its revenue performance. As consistently highlighted throughout Chapter 2, any effort aimed at answering such research objectives has to incorporate the consumer choice behavior in the optimization routines embedded in the RM systems. To accomplish this, however, customer-centric data that allows, at a minimum, to quantitatively understand the tradeoffs customers make when they face a choice situation needs to be available.

The customer-centric data should provide any interested party with at least two types of information. First, and typically, major service providers do collect and own such information due to their advanced IT infrastructures, the purchase transaction level data gives the analysts all the details associated with consumed transactions. Among other things, the analyst knows at least at a confirmation number level who purchased a specific product, when the transaction was originated and, equally important, when it consumed, what characteristics the purchased product had and whether the customer had

been enrolled in the service provider's Rewards Program, if any, at the time she initiated the transaction. Second, to assist in the development of analytic tools that explain and replicate the customer purchasing behavior, time of booking availability information is also needed. In particular, the products a customer is shown with when she makes a purchase decision together with their intrinsic attributes such as price and the corresponding purchasing restrictions need to supplement the data that comes from other sources. The combined data helps develop rational models of consumer purchasing behavior that can be incorporated in the RM decision mechanisms that lead to the implementation of the preferred capacity control policies. A unique methodology that allows us to collect customer-centric data of the type discussed above from a major competitor in the hospitality industry is presented and discussed in detail in Chapter 4. To the best of our knowledge, this data collection methodology is the first successful attempt in the industries that employ RM systems intended to provide the study sponsor with a sound, systematic way of collecting customer-centric data.

As mentioned above, the purchase transaction level data enriched with the time of booking availability information enables a service provider to gain insights into the purchasing behavior of its customers. Yet, the incompleteness of the purchase transaction level data, that is, the presence in the data of only those customers who requested and purchased the service, impedes the direct use of the random utility theory that governs the derivation of the traditional discrete choice models. To overcome the missing data limitations and describe customers' purchasing behavior as accurately as possible, estimation techniques that account for the random character of the no purchase decision periods are employed instead. In particular, the Expectation-Maximization algorithm which successively adjusts the importance given in the log likelihood function to the no purchase decision periods is used in an attempt to comprehensively explain the customer purchasing behavior. In such a framework, what differentiates the current approach from the traditional decision making methodology is that the customer product

preference as well as the no purchasing tendencies can simultaneously be assessed by iteratively replacing the missing information with its expected value.

The models that describe the customer purchasing behavior set the grounds for the practical development of the choice-based RM systems. Due to the fact that the decision of a customer is now associated with an unambiguous probability of purchase, optimal capacity control policies that take into account customers' behavior can be explored and replace the widely used sub-optimal heuristic capacity controls. In this case, the focus shifts from when and what product classes the service provider needs to close or open to stay efficient to what products it should display to maximize its revenues and incentivize customers to purchase. To accomplish this task, a multi-stage decision process that recursively quantifies the total optimal expected revenues associated with all possible levels of the remaining capacity is employed. The product sets that result in these optimal values determine the optimal display policy that the service provider should follow. Given that they are the outcome of a behavioral process, not only optimally contribute these product sets to the service provider's expected revenues but also they theoretically present the customers with the product offerings that incentivize them the most to consider purchasing a product. The flow chart that describes each stage of the development of a choice-based RM system is depicted in the upper part of Figure 5 under the Development Phase heading.

The resulting preferred control policy, while mathematically optimal, cannot justify the development and deployment of a choice-based RM system unless its revenue performance under normal operating conditions is proven to outperform the performance of the RM systems currently in use. As a result, a methodology that supports the empirical comparison of the revenue performance of such competing systems is devised and briefly explained next. Essentially, based on the customer-centric data collected from the sponsor of this study, the capacity control policies associated with the competing RM systems are independently computed. Arrival streams, which follow the

observed demand patterns for offered product classes, are simulated and, subsequently, loaded onto the systems managed by these capacity controls (for more details, see the heading Revenue Performance Evaluation Phase in Figure 5). The average revenue performance associated with the use of each control policy is finally quantified and used to make recommendations with respect to whether the development of the choice-based RM systems is warranted. In this context, several issues related to the generalizability of the findings to other industries are also discussed. The system evaluation process is schematically depicted in the lower part of Figure 5 under the System Evaluation Phase heading.

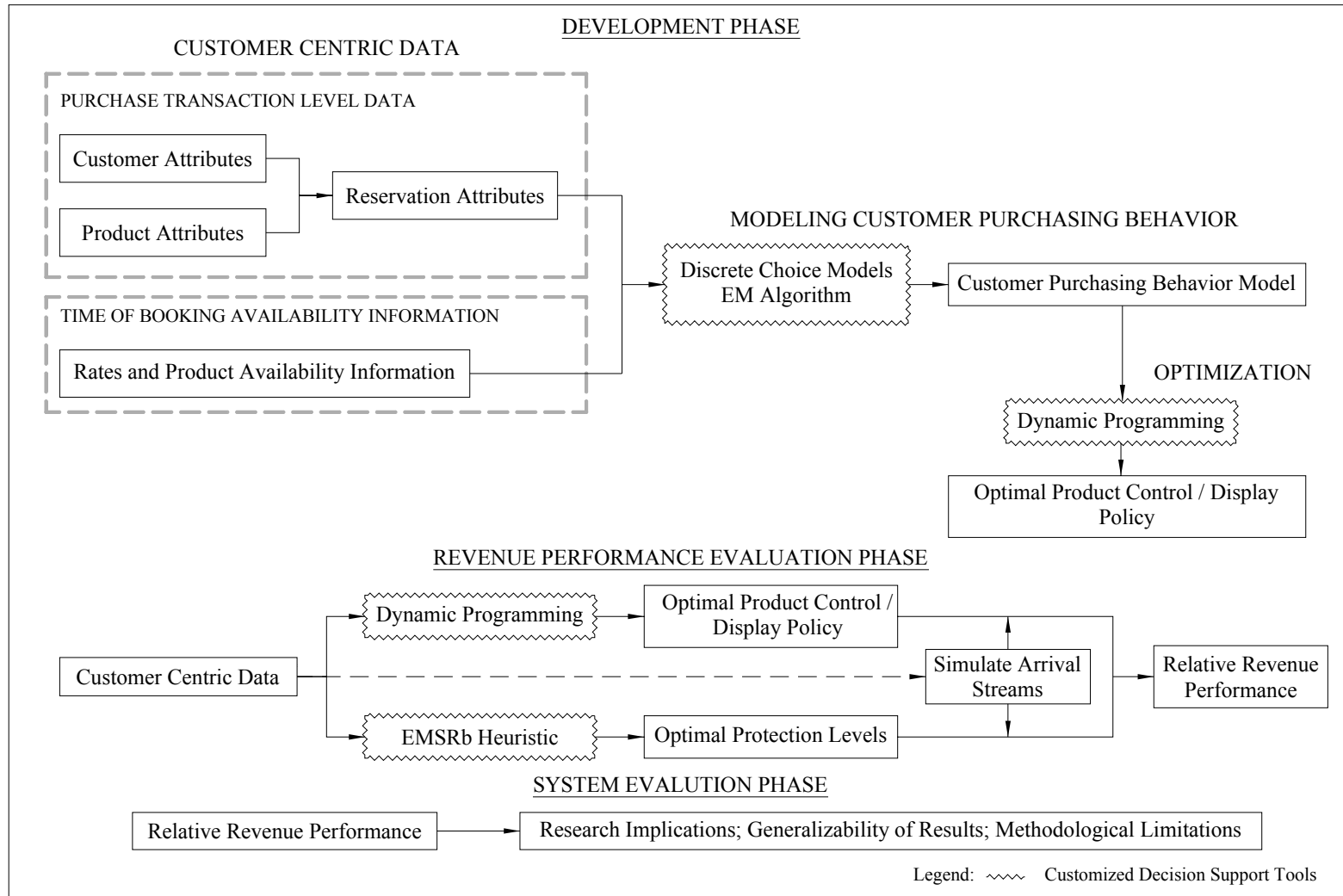


Figure 5 Refinement of Conceptual Model

3.4. Modeling Customer Purchasing Behavior

3.4.1. Multinomial Logit Model

In various settings that relate to economics, psychology, marketing, transportation, etc., discrete choice models are used in an attempt to explore and explain how generic decision makers select a preferred alternative from a set of competing alternatives.

Within a typical discrete choice framework, the alternatives available to a decision maker in a choice situation need to be mutually exclusive, exhaustive, and finite. This means that the decision maker is capable of differentiating among alternatives and, based on her preference, she always selects a single alternative from the available set. In addition, this implies that the decision maker faces a choice set that consists of all possible alternatives and, subsequently, that the cardinal of this set is necessarily finite.

Often, discrete choice models are derived based on the assumption that the decision maker is a rational being who makes choices in such a way that her utility is maximized. Implicitly, this suggests that the decision maker associates a utility value to all alternatives she is considering and selects the one that provides her with the highest such value. In this context, the behavioral model that an analyst/researcher tries to describe refers, therefore, to the decision maker's mental actions that result in her choosing alternative i over all other available alternatives j . Mathematically, this reduces to the analyst explaining the decision maker's choice as the result of a consciously maximization process over the utilities U_j of all available alternatives j . In particular, the decision maker, referred to hereafter as n in all subsequent mathematical formulas, is assumed to choose alternative i iff the utility U_{ni} is greater than the utilities U_{nj} of all alternatives $j \neq i$.

While precisely known by the decision maker at the time of her selection, the utilities U_{nj} are unknown to the analyst. As a result, the latter is forced to express these

utilities as functions of various alternative specific, decision maker specific and/or environmental specific attributes which are assumed to influence the choice and can be observed and appropriately measured. The intention here is to explain as much of the utilities U_{nj} as possible through the use of the observed attributes and then incorporate the influence of all other factors that affect the choice but are not included in the utility formulation within a random, stochastic and unobserved component. To this end, the analyst decomposes the utility U_{nj} of the j th alternative as $U_{nj} = V_{nj} + \varepsilon_{nj}$, where V_{nj} and ε_{nj} are the observed and unobserved, respectively, components of the utility U_{nj} . In this formulation, the relationship between V_{nj} and ε_{nj} assumes independency and additivity, two properties generally accepted in the discrete choice modeling field. Similarly, the observed component of the utility V_{nj} is typically considered to be linear in parameters and, therefore, it may be written as $V_{nj} = \beta' \cdot x_{nj}$ where x_{nj} is a vector of observed attributes that identify alternative j and β is a vector of parameters that needs to be estimated. The vector of parameters β quantifies the relative contribution of each attribute to the observed component of the utility.

The utility decomposition provided in the previous paragraph together with the assumed behavioral choice rule facilitates the development of a choice model that can be employed to compute the vector of parameters β . However, given that the analyst is unaware of all the factors that lead to the decision maker's choice, that is, given that ε_{nj} is a random component that takes on unknown values, the researcher cannot make absolute statements relative to the decision maker's choice. Rather, she can explain the choice only up to a probability that a current alternative is going to be selected. Thus, the probability the decision maker n chooses alternative i can be expressed as follows:

$$\begin{aligned}
 P_{ni} &= P(U_{ni} > U_{nj}; \forall j, j \neq i) = P(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}; \forall j, j \neq i) \\
 P_{ni} &= P(V_{ni} - V_{nj} > \varepsilon_{nj} - \varepsilon_{ni}; \forall j, j \neq i) = P(\varepsilon_{nj} < V_{ni} - V_{nj} + \varepsilon_{ni}; \forall j, j \neq i). \quad (1)
 \end{aligned}$$

Evaluating over all possible values l_r of ε_{ni} , the expression of P_{nj} given in equation (1) becomes:

$$P_{ni} = \sum_{r=1}^R \left[P(\varepsilon_{ni} = l_r) \cdot P(\varepsilon_{nj} < V_{ni} - V_{nj} + l_r; \forall j, j \neq i) \right]. \quad (2)$$

or, alternatively:

$$P_{ni} = \int_{-\infty}^{+\infty} \left[P(\varepsilon_{ni} = l_r) \cdot P(\varepsilon_{nj} < V_{ni} - V_{nj} + l_r; \forall j, j \neq i) \right]. \quad (3)$$

Equation (3) lays the foundation for the derivation of specific discrete choice models. However, to fully operationalize these models, additional assumptions need to be made. Specifically, the unobserved components ε_{nj} of the decision maker's utilities U_{nj} , while random in nature, have to be associated with well specified statistical distributions that control their dispersion in the utility space. Amongst all possible distributions that can be employed, the Extreme Value Type 1 and the normal distributions are by far the most popular. In addition, for each decision maker n , the relationship amongst the unobserved components ε_{nj} of alternatives j needs to be also specified. In various settings, researchers investigate multiple specifications that go from the ε_{nj} s being independent and identically distributed until them being correlated as well as unidentically distributed. Finally, for identification purposes, further constraints on the unobserved components of the utilities need to be introduced. For example, since both the absolute magnitudes and the scales of the utilities are irrelevant, at least the variance of one of the ε_{nj} s is mandatory constrained to an arbitrarily chosen value, typically equal to 1.

The assumptions the analyst makes with respect to the issues detailed above lead to the derivation of different discrete choice models. The most important of all, that is, the Multinomial Logit Model (MNL) McFadden [58] which is used later on to determine the optimal product display policy at a service provider is examined next. The MNL model requires the unobserved components of the utility to follow an Extreme Value Type 1 distribution, be independent and identically distributed and have the variances set

to 1. With these assumptions in place, the expression depicted in equation (3) for the probability of decision maker n selecting alternative i simplifies and, after some algebraic manipulations, becomes:

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_{j=1}^J e^{V_{nj}}} = \frac{e^{\beta' \cdot x_{ni}}}{\sum_{j=1}^J e^{\beta' \cdot x_{nj}}} \quad (4)$$

where: P_{ni} is the probability of decision maker n choosing alternative i ,
 J provides the cardinal of the decision maker's choice set,
 V_{ni} and V_{nj} are the observed components of the utilities of alternatives i and j ,
 x_{ni} and x_{nj} are vectors of observed attributes that identify alternatives i and j ,
 β is a vector of unknown parameters,
 e refers in this context to the exponential function.

The probability expression given in equation (4) establishes a relationship amongst the observed attributes x_n , the unknown parameters β and the probabilistic outcome P_{ni} . The intention here is to estimate the parameters β in such a way that the probabilistic outcomes P_{ni} match as closely as possible the decision makers' observed choices. Stated otherwise, through several iterations that alter the values of the parameters β , the aim is to maximize the likelihood that the observed choices are correctly identified by the model across all decision makers in the data sample. Thus, we want to determine the vector of parameters β that maximizes the likelihood function

$$L = \prod_{n=1}^N \left[\prod_{i=1}^J (P_{ni}^{\delta_{ni}}) \right] \quad (5)$$

where: N provides the number of decision makers in the sample,
 J gives the cardinal of the decision maker n 's choice set,
 P_{ni} is the probability of decision maker n choosing alternative i ,
 δ_{ni} is a dummy variable equal to 1 if alternative i is the observed choice of decision maker n , 0 otherwise.

While easy to compute for less complicated likelihood functions, more involved functions may run in numeric instabilities and not provide the analyst with a correct set on β parameters. For this reason, in most applications that involve discrete choice models it is more common to maximize the log of the likelihood function instead of the function itself. For the MNL case, the log of the likelihood function depicted in equation (5) takes on the following expression:

$$LL = \log(L) = \log \left\{ \prod_{n=1}^N \left[\prod_{i=1}^J (P_{ni}^{\delta_{ni}}) \right] \right\} = \sum_{n=1}^N \left[\log \left(\prod_{i=1}^J (P_{ni}^{\delta_{ni}}) \right) \right] = \sum_{n=1}^N \sum_{i=1}^J \left[\log(P_{ni}^{\delta_{ni}}) \right]$$

or, equivalently,

$$LL = \sum_{n=1}^N \sum_{i=1}^J \delta_{ni} \cdot \log(P_{ni}). \quad (6)$$

To illustrate how the observed attributes and choices are used to compute the preferred set of parameters β , let's consider the hypothetical example of a hotel's customers who can select their rooms of choice from a maximum of four possible room types: *King Room - Advance Purchase*, *Queen Room - Advance Purchase*, *King Room - Rack Rate*, and *Queen Room - Rack Rate*. In this case, the *Advance Purchase* rooms are priced below the *Rack Rate* products but are only available up to 7 days in advance of the arrival date. In addition, if a customer requests an *Advance Purchase* room, she is required to pay a deposit which needs to cover the entire length of her stay and which is not refundable. To make the exposition clearer, we call the *Advance Purchase* rooms restricted strictly because several booking restrictions limit their use. Lastly, it should be mentioned that the room prices are set dynamically during the booking horizon based on the equilibrium of room supply and demand and, therefore, it should not come as a surprise the fact that each customer faces a different set of time of booking room prices. For convenience, the observed room attributes and choices for a subsample of three customers and a given arrival date is provided in the upper part of Table 2.

Table 2 MNL Model - Input Data

Customer	Room Type	Alternative	Chosen	ASC2	ASC3	ASC4	Price	Restricted
1	King Room - Advance Purchase	1	1	0	0	0	\$200.0	1
1	Queen Room - Advance Purchase	2	0	1	0	0	\$175.0	1
1	King Room - Rack Rate	3	0	0	1	0	\$275.0	0
1	Queen Room - Rack Rate	4	0	0	0	1	\$250.0	0
2	Queen Room - Advance Purchase	2	0	1	0	0	\$190.0	1
2	King Room - Rack Rate	3	1	0	1	0	\$290.0	0
2	Queen Room - Rack Rate	4	0	0	0	1	\$265.0	0
3	King Room - Rack Rate	3	0	0	1	0	\$280.0	0
3	Queen Room - Rack Rate	4	1	0	0	1	\$255.0	0
...								
Parameters estimates β (current values)				-0.70	0.90	0.40	-0.01	-0.75

For simplicity, let's suppose that the customer choice at this hotel is determined by the price and the restrictions, if any, associated with the room types. In this context, it is then reasonable to write the utility that a customer n obtains from purchasing the i th room type as:

$$U_{ni} = V_{ni} + \varepsilon = \beta' \cdot x_{ni} + \varepsilon = \underbrace{ASC_{ni} \cdot 1 + \beta_P \cdot Price_{ni} + \beta_R \cdot Restr_{ni}}_{\beta' \cdot x_{ni}} + \varepsilon, \forall i, ASC_{n1} = 0. \quad (7)$$

In equation (7), the subscript ni from the unobserved utility components ε_{ni} s was removed to reflect the fact that these components are identically distributed. Similarly, to avoid proven identification problems and, therefore, to appropriately set the levels of the overall utilities, the alternative specific constant associated with *King Room - Advance Purchase* (i.e., ASC1) was left out of the model. The latter action leads to the remaining alternative specific constants representing the average effect of all factors not included in the model on the utility of the room type they accompany relative to that of *King Room - Advance Purchase*.

Using the MNL formula given in equation (4) and a set of β parameters depicted in the lower part of Table 2, the probabilities of each customer purchasing each of the room types from her time of booking choice set can be evaluated. In matrix notation, the probabilities P_{li} for the first customer can be computed as:

$$V_{li} = \begin{bmatrix} V_{11} \\ V_{12} \\ V_{13} \\ V_{14} \end{bmatrix} = \beta' \cdot x_{li} = \begin{bmatrix} 0.00 \\ -0.70 \\ 0.90 \\ 0.40 \end{bmatrix} - 0.01 \cdot \begin{bmatrix} 200.0 \\ 175.0 \\ 275.0 \\ 250.0 \end{bmatrix} - 0.75 \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -2.75 \\ -3.20 \\ -1.85 \\ -2.10 \end{bmatrix}$$

$$e^{V_{li}} = e^{[V_{11} \ V_{12} \ V_{13} \ V_{14}]} = e^{[-2.75 \ -3.20 \ -1.85 \ -2.10]} = [0.06 \ 0.04 \ 0.16 \ 0.12]$$

$$\sum_{i=1}^{J=4} e^{V_{li}} = e^{V_{11}} + \dots + e^{V_{14}} = 0.38$$

$$P_{li} = \begin{bmatrix} P_{11} \\ P_{12} \\ P_{13} \\ P_{14} \end{bmatrix} = \frac{(e^{V_{li}})'}{\sum_{i=1}^{J=4} e^{V_{li}}} = \frac{([0.06 \ 0.04 \ 0.16 \ 0.12])'}{0.38} = \begin{bmatrix} 0.17 \\ 0.11 \\ 0.41 \\ 0.32 \end{bmatrix}.$$

The contribution LL_i of the first customer to the overall log likelihood function LL is assessed using the formula depicted in equation (6). Given that this customer chose a *King Room - Advance Purchase* room type, the associated dummy variable δ_{li} takes on a non-zero value for this room type only. Thus, LL_i can be computed as:

$$LL_i = \sum_{i=1}^4 (\delta_{li} \cdot \log(P_{li})) = 1 \cdot \log(P_{11}) + 0 \cdot \log(P_{12}) + \dots + 0 \cdot \log(P_{14}) = 1 \cdot \log(0.17) = -1.79.$$

The purchase probabilities as well as the contributions to the overall likelihood function for all customers in the sample can be similarly determined. For completeness, they are provided in Table 3 below. The corresponding log likelihood function together with its first and, sometimes, its second order derivatives computed at the current level of parameters β is then used to recommend new levels for β s in such a way that, at the end of this iterative process, the LL function is maximized.

Table 3 MNL Model - Excerpt from Parameter Estimation (Intermediate Step)

Cust.	Room Type	Alt.	Chosen	V_{ni}	$\exp(V_{ni})$	$\sum(\exp(V_{ni}))$	P_{ni}	δ_{ni}	LL_n
1	King Room - Advance Purchase	1	1	-2.75	0.06	0.38	0.17	1	-1.79
1	Queen Room - Advance Purchase	2	0	-3.20	0.04		0.11	0	
1	King Room - Rack Rate	3	0	-1.85	0.16		0.41	0	
1	Queen Room - Rack Rate	4	0	-2.10	0.12		0.32	0	

Table 3 Continued

Cust.	Room Type	Alt.	Chosen	V_{ni}	$\exp(V_{ni})$	$\sum(\exp(V_{ni}))$	P_{ni}	δ_{ni}	LL_n
2	Queen Room - Advance Purchase	2	0	-3.35	0.04	0.28	0.13	0	
2	King Room - Rack Rate	3	1	-2.00	0.14		0.49	1	-0.71
2	Queen Room - Rack Rate	4	0	-2.25	0.11		0.38	0	
3	King Room - Rack Rate	3	0	-1.90	0.15	0.27	0.56	0	-0.83
3	Queen Room - Rack Rate	4	1	-2.15	0.12		0.44	1	
...									

The underlying assumptions that lead to the derivation of the closed form MNL probability formula shown in equation (4) provide a simple framework for estimating the preferred set of parameters β . In spite of this notable advantage, these assumptions - in different contexts and choice situations explored - may limit the behavioral power of the MNL model. The theoretical concepts that identify these limitations together with several illustrative examples are discussed next.

First, due to the unobserved components of the utility being considered independent and identically distributed, the MNL model does not account for random taste variation. To explain this, let's go back to the example whose data is described in Table 2 and consider that the value customers place on the room price varies in the customer population and depends on the multiplicative inverse of the customer's income and the comfort the customer associates with the paid price. Thus, the parameter β_p associated with the room price becomes specific to each individual customer n and can be expressed as

$$\beta_{p,n} = \frac{\alpha}{Income_n} + \gamma \cdot Comfort_n.$$

From a behavioral perspective, the assumed relationship suggests that lower income customers are more price sensitive than the higher income ones and that a higher perceived comfort adds more to the utility of all room types. Substituting this relationship into equation (7) results in

$$U_{ni} = ASC_{ni} + \left(\frac{\alpha}{Income_n} + \gamma \cdot Comfort_n \right) \cdot Price_{ni} + \beta_R \cdot Restrictions_{ni} + \varepsilon_{ni},$$

or, alternatively:

$$U_{ni} = ASC_{ni} + \alpha \cdot \frac{Price_{ni}}{Income_n} + \gamma \cdot Comfort_n \cdot Price_{ni} + \beta_R \cdot Restrictions_{ni} + \varepsilon_{ni}. \quad (8)$$

Under the assumption that ε_{ni} s are independent and identically distributed, if all variables in equation (8) are observed then the MNL model can accurately describe the decision making process customers go through when they buy lodging services.

Building on the above ideas, let's suppose now that the analyst is well aware of the fact that customer's perceived comfort influences the room choice but she can neither observe it directly nor recreate it indirectly, hence she treats it as a random variable. Given the circumstances, the analyst is forced to leave comfort out of the model and, for the same problem, specify a utility function that looks like:

$$U_{ni} = ASC_{ni} + \alpha \cdot \frac{Price_{ni}}{Income_n} + \beta_R \cdot Restrictions_{ni} + \varepsilon_{ni}^N, \quad (9)$$

where $\varepsilon_{ni}^N = \gamma \cdot Comfort_n \cdot Price_{ni} + \varepsilon_{ni}$. In this case, since the unobserved components ε_{ni}^N s are not independent (i.e., customer's perceived comfort enters the utility of all room types and, therefore, introduces correlation among room types) and identically distributed (i.e., the variance of ε_{ni}^N depends on the characteristics of room type i), the MNL model becomes an unrealistic modeling mechanism. Based on this example, if the intention is to represent customers' random preference for products, more flexible choice models need to be employed since the MNL model constitutes a misspecification.

The second limitation refers to the MNL model imposing restrictive substitution patterns among alternatives. To provide the context for the discussion, let's consider that the analyst, working with the data described in Table 2, is able to estimate a reasonably

robust MNL model that predicts the probabilities for an average customer choosing each of the available room types as those provided in the appropriate column of Table 4 below.

Table 4 MNL Model - Average Customer's Purchasing Probabilities (Example)

Room Type	ASC2	ASC3	ASC4	Price	Restricted Prob.	Adj. Prob.
King Room - Advance Purchase	0	0	0	\$200.0	1	0.30
Queen Room - Advance Purchase	1	0	0	\$175.0	1	0.32
King Room - Rack Rate	0	1	0	\$275.0	0	0.17
Queen Room - Rack Rate	0	0	1	\$250.0	0	0.21

Armed with this information, the analyst may want to investigate how these probabilities redistribute when certain room types are removed or added from or to the customer choice sets. To make the exposition concrete, let's suppose that the analyst intends to compute the redistributed probabilities when *Queen Room - Advance Purchase* is removed from the choice sets. With a good understanding of the hospitality industry, the analyst expects that customers who would have previously purchased the removed room type would primarily migrate towards the alternatives that are similarly priced or, instead, would choose not to buy. As a result, she anticipates that from the customers who decide to purchase, many will try to book a *King Room - Advance Purchase*. For this particular room type, a resulting estimate for the probability of purchase in the 0.48 - 0.52 range would represent an outcome that would be consistent with the analyst's expectations.

The use of the MNL methodology described herein allows, however, the analyst to precisely (but not necessarily correctly) quantify the shifts in the purchase probabilities due to the removal of the *Queen Room - Advance Purchase* alternative. As such, since the ratio of the MNL probabilities for two room types i and j , $P_{ni}/P_{nj} = e^{V_{ni}}/e^{V_{nj}} = e^{(V_{ni}-V_{nj})}$, depends only on the attributes of the corresponding alternatives, this ratio stays the same irrespective of what other room types are made available to customers. Mathematically, this translates into the probabilities of the remaining room types having to be adjusted by a multiplication factor of $1/(1 - P_{KR-AP})$, $P_{KR-AP} = 0.32$, to account for the imposed change. The adjusted probabilities for the remaining hotel products are shown in the last column

of Table 4. As can be easily seen, the prediction of the utilization rate for *King Room - Advance Purchase* is well below what the analyst expects and is most likely imprecise. Even if it describes a hypothetical scenario, this example highlights the mispredictions that may occur when the addition or removal of alternatives does not proportionally affect the probabilities of the previously existing or the remaining choice alternatives. For this reason, whenever other than proportional substitution patterns are hypothesized to control the probability redistribution, more flexible models such as nested logits (McFadden [59], Williamson [60]), mixed logits and/or probits (Daganzo [61], Marschak [62], Thurstone [63]) need to be employed.

The last limitation of the MNL models concerns the way they handle choice situations where multiple choices are observed for the same decision maker over time. In the hospitality industry, for example, if the analyst is interested in describing the dynamics of room type choice, she may survey a sample of representative customers for the details of their hotel stays consumed in the last, say, two years. Evidently, based on their travel patterns, some of the customers may show multiple entries in the survey database. In such a context, the theoretical findings presented so far are still applicable. Thus, if the unobserved components of the utility of customers with multiple choices are independent of each other over time, the MNL model can be used to research the analyst's interests. In this case, the choices made at different times by the same customer are processed independently as they were coming from multiple unrelated customers. However, if the unobserved utility components of the multiple choices are related over time (e.g., a negative hotel experience, an event unobserved by the analyst, tends to influence the customer choice at all subsequent times), the MNL model is once again a misspecification. To cope with such choice situations, the analyst can try to effectively quantify the correlated unobserved factors and place them in the observed part of the utility, or, alternatively, she can employ more flexible discrete choice models such as mixed logits and probits.

3.4.2. Expectation Maximization Algorithm

In the hospitality industry, customers make choices at various levels. At one end of the spectrum, they need to decide based on the purpose of the travel the destination of their trips and identify the means of how to get there. At the other, based on their personal preference and travel constraints, they have to choose among several competing hotel services and products. In between these extremes, customers are typically required to select at the preferred destination their hotel of choice.

Focusing on the second of the extreme cases, it is generally accepted that the task a customer has to undertake is not straightforward. Given the overwhelming product variety, the customer has to devote a substantial amount of time to balance the price and the characteristics of the available products with the flexibility required by her travel plans and schedule. Due to the fact that a matching is not guaranteed, or, simply because the customer considers that the searching costs exceed the benefits associated with a purchase, it may happen that initiated product requests result in nothing being purchased. For example, internet conversion rate estimates in the 5% - 11% range are commonly reported in the hospitality industry (Petrova [64], Smith, et al. [65]). Since the customers who initiate but do not terminate a product purchase do not directly contribute to a company's profit, most hospitality operators do not keep detailed records for them. If such records existed, that is, if the selected products, if any, and the time of booking choice sets for all customers who solicited hotel products and *purchased* or *did not purchase* a preferred one were available, the MNL methodology as described in Sub-section 3.4.1 could be used to investigate customer purchasing behavior in a hotel environment. Lack of such information, however, restricts the applicability of this methodology and forces the analyst to approach solving her problem differently. In particular, the analyst faces a missing data problem whose set-up and solution are detailed next.

In an attempt to illustrate the source of problems the analyst encounters let's take a look first at how the reservations are typically made at a given hotel. For a specific arrival date, customers can request price quotes and, subsequently, book a product at any point in time during the booking horizon. Phone calls made directly to the hotel reception desk or to a Central Reservation Office, inquiries about price and room availability that come from travel agents or customers' own attempts to book their hotel services on line, all qualify in this context as requested quotes. To help the exposition, a possible stream of requests is depicted in the upper panel of Figure 6.

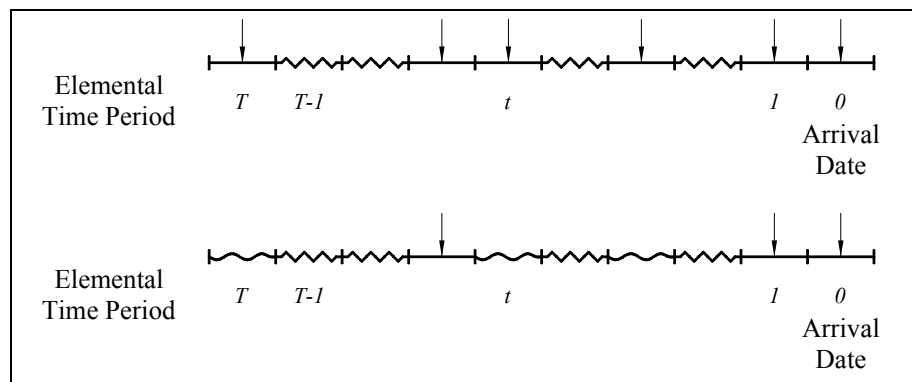


Figure 6 Customer Arriving Process - Example

Since the number of requests for hotel products is finite, the booking horizon can be uniformly discretized in such a way that any given arrival coincides with a single elemental time period. For illustration purposes, the elemental time periods where no product requests are observed are marked with zigzag lines. Moving forward, let's suppose that of all observed arrivals only some are associated with a product purchase. This assumption intends to replicate the well known fact that many of the requests are phony and do not result in consumed room reservations. For our exercise, the arrivals that do consume are shown in the lower panel of Figure 6. These arrivals are typically represented in the purchase transaction level data that most hospitality service providers collect and store. To facilitate the explanation and differentiate them from the consumed arrivals, the elemental time units that have a request but do not result in a purchase are marked with curved lines.

Within the outlined framework, the conflict arises when the analyst tries to model the choice situation presented in the upper panel of Figure 6 with the data depicted in its lower panel. The incomplete information, that is, the inability of the analyst to accurately discriminate between time periods with no requests at all (i.e., the zigzag lines) and those with requests but not observed purchases (i.e., the curved lines), makes the approach described in Sub-section 3.4.1 inapplicable to this problem. A modified version of the MNL methodology that accounts for the dual cause of a no purchase time period (i.e., the zigzag and curved lines) is introduced next. This methodology, which involves the application of the expectation - maximization (EM) techniques, was originally used by Talluri and van Ryzin [14] in an airline environment and represents an extension of the fundamental work of Dempster, et al. [66].

To begin with, the intention is to analytically describe the purchasing behavior of customers who reserve hotel products through several distribution channels. Each potential customer is assumed to select her preferred room type from a set of available such products. More over, the composition of the choice sets is assumed to vary during the booking horizon, with room types being added or removed as a function of the forecast demand, the remaining room capacity, and the remaining time until the arrival date. In addition, customers are considered to always have the alternative of buying none of the offered products among the alternatives in their choice sets. Consistent with a previous remark, it is also assumed that the booking horizon can be discretized so finely that an elemental time period can be associated with at most one arrival. For simplicity, λ , the arrival rate of customers, is assumed to stay the same over the entire booking horizon, that is, the arrivals are assumed to follow a homogeneous Poisson process. In addition, customers are assumed to make decisions according to the random utility theory, i.e., when faced with multiple purchase options, they choose the room type that maximizes their utility. For the purpose of this discussion, the probability of selecting one room type among a set of mutually exclusive and collectively exhaustive room types

is considered to be given by a multinomial logit model. The arrival rate λ and the MNL parameters β are unknown and need to be determined based on the available purchase transaction level data. Following then the notations in Talluri and van Ryzin [14], we use D , P , and $\bar{P} = D - P$ to denote the entire set of elemental time periods, the set of periods in which a purchase is observed, and the set of periods for which no-purchase transactions are reported, respectively (e.g., in the lower panel of Figure 6), all T time periods enter D , the periods represented by the straight lines determine P , and the periods marked with curved and zigzag lines identify \bar{P}). Furthermore, to facilitate the exposition, $j(t)$ is considered to represent the product the customer arriving in elemental time period $t \in P$ purchases from the product choice set $S(t)$ she is offered for consideration. With the notations formally defined, the complete likelihood function for the entire booking horizon becomes:

$$L = \prod_{t \in P} \left\{ \left[\lambda \cdot P_{j(t)}(x, \beta, S(t)) \right]^{a(t)} \cdot (1 - \lambda)^{(1-a(t))} \right\} \cdot \prod_{t \in \bar{P}} \left\{ \left[\lambda \cdot P_0(x, \beta, S(t)) \right]^{a(t)} \cdot (1 - \lambda)^{(1-a(t))} \right\}, \quad (10)$$

where: $P_{j(t)}(x, \beta, S(t))$ is the purchasing probability associated with the purchase period t and choice made $j(t)$,

$P_0(x, \beta, S(t))$ is the no-purchasing probability associated with the no-purchase period t ,

x is a vector of known attributes (e.g., price, product restrictions) that describes each product in the consideration set,

$a(t)$ is an indicator variable equal to 1 if a customer's request for a product arrives in period t , 0 otherwise.

In equation (10), $a(t)$ in the first product is always equal to 1 since an observed purchase necessarily requires an arrival. As a result, the above likelihood function can be re-written as:

$$L = \prod_{t \in P} [\lambda \cdot P_{j(t)}(x, \beta, S(t))] \cdot \prod_{t \in \bar{P}} \left\{ [\lambda \cdot P_0(x, \beta, S(t))]^{a(t)} \cdot (1-\lambda)^{(1-a(t))} \right\}. \quad (11)$$

In the expression depicted in equation (11), the product over all periods $t \in P$ computes the likelihood that the observed purchase outcomes are correctly replicated by the behavioral model employed. This likelihood is then adjusted through the use of the second product to account for the influence of the arrival - no purchase mechanism that characterizes the time periods in \bar{P} . In particular, in the second product, it is acknowledged that a no purchase can be caused by a customer requesting a product and deciding not to buy (i.e., $\lambda \cdot P_0(x, \beta, S(t))$), or, by a customer not requesting service at all (i.e., $(1-\lambda)$). Since these circumstances exclude each other, the presence of the exponents $a(t)$ and $(1-a(t))$, respectively, guarantees the independence of the two events.

Taking the logarithm of L in equation (11) we obtain the complete log likelihood function for the entire booking horizon, which can be expressed as follows:

$$\begin{aligned} LL = \ln(L) &= \ln \prod_{t \in P} [\lambda \cdot P_{j(t)}(x, \beta, S(t))] + \\ &\quad + \ln \prod_{t \in \bar{P}} \left\{ [\lambda \cdot P_0(x, \beta, S(t))]^{a(t)} \cdot (1-\lambda)^{(1-a(t))} \right\} \\ &= \sum_{t \in P} \left(\ln [\lambda \cdot P_{j(t)}(x, \beta, S(t))] \right) + \\ &\quad + \sum_{t \in \bar{P}} \left(\ln \left\{ [\lambda \cdot P_0(x, \beta, S(t))]^{a(t)} \cdot (1-\lambda)^{(1-a(t))} \right\} \right) \\ &= \sum_{t \in P} \left[\ln(\lambda) + \ln(P_{j(t)}(x, \beta, S(t))) \right] + \\ &\quad + \sum_{t \in \bar{P}} \left(\ln \left\{ [\lambda \cdot P_0(x, \beta, S(t))]^{a(t)} \right\} + \ln \left[(1-\lambda)^{(1-a(t))} \right] \right) \\ &= \sum_{t \in P} \left[\ln(\lambda) + \ln(P_{j(t)}(x, \beta, S(t))) \right] + \\ &\quad + \sum_{t \in \bar{P}} \left\{ a(t) \cdot \left[\ln(\lambda) + \ln(P_0(x, \beta, S(t))) \right] + (1-a(t)) \cdot \ln(1-\lambda) \right\}. \quad (12) \end{aligned}$$

In the log likelihood function depicted in equation (12), besides parameters λ and β , the indicator variables $a(t)$ for time periods t with no observed purchases are also unknown. This situation is best explained by returning to the lower panel of Figure 6 where the analyst is given no means to differentiate between the time periods marked by curved and zigzag lines. In particular, based on the purchase transaction level data that is available to her, she cannot unmistakably tell which of the time periods with no observed purchases are associated with immaterialized product requests (i.e., the identification of the curved lines is not possible). The inability to unequivocally discriminate between these distinct time periods constitutes the reason why the maximum likelihood estimation technique detailed in Sub-section 3.4.1 cannot be directly applied to process data of this type. To overcome this problem, the EM estimation technique, which requires the analyst to iteratively compute the expected values of these $a(t)$'s based on a set of current parameter estimates $\widehat{\beta}$ and $\widehat{\lambda}$, is employed instead. Since the $a(t)$'s are indicator variables that can be described by a Bernoulli distribution, their expected values take on the following expression:

$$\begin{aligned}\widehat{a}(t) &= E\left(a(t) \mid t \in \overline{P}, \widehat{\beta}, \widehat{\lambda}\right) \\ &= 1 \cdot P\left(a(t) = 1 \mid t \in \overline{P}, \widehat{\beta}, \widehat{\lambda}\right) + 0 \cdot P\left(a(t) = 0 \mid t \in \overline{P}, \widehat{\beta}, \widehat{\lambda}\right) \\ &= P\left(a(t) = 1 \mid t \in \overline{P}, \widehat{\beta}, \widehat{\lambda}\right).\end{aligned}\tag{13}$$

Applying Bayes' rule to the expression in equation (13), the expected value $\widehat{a}(t)$ becomes:

$$\widehat{a}(t) = \frac{P\left(t \in \overline{P} \mid a(t) = 1, \widehat{\beta}, \widehat{\lambda}\right) \cdot P\left(a(t) = 1 \mid \widehat{\beta}, \widehat{\lambda}\right)}{P\left(t \in \overline{P} \mid \widehat{\beta}, \widehat{\lambda}\right)},\tag{14}$$

where: $P\left(t \in \overline{P} \mid a(t) = 1, \widehat{\beta}, \widehat{\lambda}\right) = P_0\left(x, \widehat{\beta}, \widehat{\lambda}, S(t)\right)$ is the probability of a no-purchase given that a customer arrives,

$P(a(t)=1|\widehat{\beta}, \widehat{\lambda}) = \widehat{\lambda}$ is the probability of an arrival,

$P(t \in \overline{P}|\widehat{\beta}, \widehat{\lambda}) = \widehat{\lambda} \cdot P_0(x, \widehat{\beta}, \widehat{\lambda}, S(t)) + (1 - \widehat{\lambda})$ is the probability of a no-purchase.

Substituting the expression of above probabilities back in equation (14), $\widehat{a}(t)$ can be re-written as:

$$\widehat{a}(t) = \frac{\widehat{\lambda} \cdot P_0(x, \widehat{\beta}, \widehat{\lambda}, S(t))}{\widehat{\lambda} \cdot P_0(x, \widehat{\beta}, \widehat{\lambda}, S(t)) + (1 - \widehat{\lambda})}. \quad (15)$$

Substituting $\widehat{a}(t)$ given by equation (15) in equation (12), we can express the expected conditional log likelihood function for the incomplete data as:

$$\begin{aligned} E(LL|\widehat{\beta}, \widehat{\lambda}) = & \sum_{t \in P} [\ln(\lambda) + \ln(P_{j(t)}(x, \beta, S(t)))] + \\ & + \sum_{t \in \overline{P}} \left\{ \widehat{a}(t) \cdot [\ln(\lambda) + \ln(P_0(x, \beta, S(t)))] + (1 - \widehat{a}(t)) \cdot \ln(1 - \lambda) \right\}. \end{aligned} \quad (16)$$

Furthermore, given that $E(LL|\widehat{\beta}, \widehat{\lambda})$ is separable in parameters, the values of λ and β that maximize the log likelihood function in equation (16) can be independently computed. Specifically, taking the derivative of $E(LL|\widehat{\beta}, \widehat{\lambda})$ with respect to λ and imposing it to be equal to zero, we obtain:

$$\begin{aligned} \frac{d}{d\lambda} (E(LL|\widehat{\beta}, \widehat{\lambda})) &= \frac{d}{d\lambda} \left(\sum_{t \in P} \ln(\lambda) \right) + \frac{d}{d\lambda} \left(\sum_{t \in P} [\widehat{a}(t) \cdot \ln(\lambda)] \right) + \\ &+ \frac{d}{d\lambda} \left(\sum_{t \in \overline{P}} [(1 - \widehat{a}(t)) \cdot \ln(1 - \lambda)] \right) \\ &= \sum_{t \in P} \left(\frac{1}{\lambda} \right) + \sum_{t \in \overline{P}} \left(\frac{\widehat{a}(t)}{\lambda} \right) + \sum_{t \in \overline{P}} \left((1 - \widehat{a}(t)) \cdot \frac{d}{d\lambda} (\ln(1 - \lambda)) \right) \\ &= \sum_{t \in P} \left(\frac{1}{\lambda} \right) + \sum_{t \in \overline{P}} \left(\frac{\widehat{a}(t)}{\lambda} \right) + \sum_{t \in \overline{P}} \left((1 - \widehat{a}(t)) \cdot \frac{1}{1 - \lambda} \cdot (-1) \right) \end{aligned}$$

$$\begin{aligned}
&= \frac{|P|}{\lambda} + \frac{\sum_{t \in P} (\hat{a}(t))}{\lambda} + \frac{\sum_{t \in \bar{P}} (\hat{a}(t) - 1)}{1 - \lambda} \\
&= \frac{|P| + \sum_{t \in P} (\hat{a}(t)) - \lambda \cdot (|P| + |\bar{P}|)}{\lambda \cdot (1 - \lambda)} = 0, \text{ or, equivalently,} \\
&\lambda^* = \frac{|P| + \sum_{t \in P} (\hat{a}(t))}{|P| + |\bar{P}|}. \tag{17}
\end{aligned}$$

The derivation that leads to the expression depicted in equation (17) does not guarantee that the conditional log likelihood function $E(LL|\widehat{\beta}, \widehat{\lambda})$ is maximized with respect to λ at λ^* . To prove, however, that this is the case, the second derivative of $E(LL|\widehat{\beta}, \widehat{\lambda})$ with respect to λ computed at λ^* is shown to be negative. As such, if λ is substituted with λ^* in the expression of the second derivative

$$\begin{aligned}
&\frac{d^2}{d\lambda^2} \left(E(LL|\widehat{\beta}, \widehat{\lambda}) \right) = \frac{d}{d\lambda} \left(\frac{|P| + \sum_{t \in P} (\hat{a}(t)) - \lambda \cdot (|P| + |\bar{P}|)}{\lambda \cdot (1 - \lambda)} \right) \\
&= \frac{[\lambda \cdot (1 - \lambda)] \cdot \frac{d}{d\lambda} \left(|P| + \sum_{t \in P} (\hat{a}(t)) - \lambda \cdot (|P| + |\bar{P}|) \right) - \left[|P| + \sum_{t \in P} (\hat{a}(t)) - \lambda \cdot (|P| + |\bar{P}|) \right] \cdot \frac{d}{d\lambda} (\lambda \cdot (1 - \lambda))}{[\lambda \cdot (1 - \lambda)]^2} \\
&= \frac{-1 \cdot [\lambda \cdot (1 - \lambda)] \cdot (|P| + |\bar{P}|) - \left[|P| + \sum_{t \in P} (\hat{a}(t)) - \lambda \cdot (|P| + |\bar{P}|) \right] \cdot (1 - 2 \cdot \lambda)}{[\lambda \cdot (1 - \lambda)]^2}
\end{aligned}$$

an entity whose value is always negative is obtained. For completeness, its parametric expression is provided in equation (18) below:

$$\begin{aligned}
\frac{d^2}{d\lambda^2} \left(E \left(LL | \widehat{\beta}, \widehat{\lambda} \right)_{\lambda^*} \right) &= - \frac{|P| + |\overline{P}|}{\lambda^* \cdot (1 - \lambda^*)} = - \frac{|P| + |\overline{P}|}{\frac{|P| + \sum_{t \in \overline{P}} (\widehat{a}(t))}{|P| + |\overline{P}|} \cdot \left(1 - \frac{|P| + \sum_{t \in \overline{P}} (\widehat{a}(t))}{|P| + |\overline{P}|} \right)} \\
&= - \frac{(|P| + |\overline{P}|)^2}{\left[|P| + \sum_{t \in \overline{P}} (\widehat{a}(t)) \right] \cdot \left[|\overline{P}| - \sum_{t \in \overline{P}} (\widehat{a}(t)) \right]}, \quad |\overline{P}| > \sum_{t \in \overline{P}} (\widehat{a}(t)). \quad (18)
\end{aligned}$$

Similar reasoning applies to finding the parameter estimates β^* that maximize the component of $E(LL|\widehat{\beta}, \widehat{\lambda})$ that depends on β only. Here, we need to compute β^* by solving:

$$\max \left(\sum_{t \in P} \left[1 \cdot \ln(P_{j(t)}(x, \beta, S(t))) \right] + \sum_{t \in \overline{P}} \left[\widehat{a}(t) \cdot \ln(P_0(x, \beta, S(t))) \right] \right). \quad (19)$$

Since computing the optimal values for β duplicates the methodology presented in Subsection 3.4.1 no other attempts are made here to discuss this topic.

Parameter estimates β^* and λ^* obtained from equations (17) and (19) are used next to verify some predefined convergence criteria and, if these are not met, the procedure described in (13) - (19) is repeated. A summary of the entire algorithm, as provided in Talluri and van Ryzin [14], is given below:

Step 0:	Initialization step Initialize $\widehat{\lambda}$ and $\widehat{\beta}$.
Step 1:	Expectation step Given z , $\widehat{\beta}$, $\widehat{\lambda}$, and $S(t)$ use equation (15) to compute $\widehat{a}(t)$ for all periods t in \overline{P} .
Step 2:	Maximization step Compute λ^* using equation (17). Compute β^* using equation (19).
Step 3:	Convergence check If <i>convergence criteria</i> then λ^* and β^* are optimal; else $\widehat{\lambda} \leftarrow \lambda^*$, $\widehat{\beta} \leftarrow \beta^*$ and redo steps 1-3.

The EM algorithm as discussed above may seem to be a sophisticated alternative to the straightforward MNL methodology presented in Sub-section 3.4.1. Such an opinion, if formed primarily on the basis of the method's involved analytical requirements, may, however, be misleading. The EM algorithm makes use of all advantages, theoretical and practical, that come with the MNL model but repeatedly adjusts its optimal solution to account for the random cause of the no purchase events. To practically illustrate how the missing information is incorporated in the preferred solution, let's revisit the hypothetical example described in Table 2 and adapt it to better support our intentions. As such, as shown in Table 5, the booking horizon is assumed to consist of four elemental time periods T_i , the first and the fourth of which are associated with consumed reservations. In this case, product availability is given by the presence or the absence of the product price in the corresponding time period fields. During the fourth elemental time period T_4 , for example, only two of all possible products are available. In addition, for the periods associated with a purchase, the highlighted price field identifies the chosen product. To simplify the exposition, all other variables that enter this model carry the meanings that were previously discussed.

Table 5 EM Algorithm - Problem Set Up

Alternative	Elemental Time Period ^{a)}				ASC2	ASC3	ASC4	ASC5	Rest.
	T_4	T_3	T_2	T_1					
King Room - Advance Purchase	\$200.0	\$210.0	-	-	0	0	0	0	1
Queen Room - Advance Purchase	\$175.0	-	-	-	1	0	0	0	1
King Room - Rack Rate	\$275.0	\$290.0	\$300.0	\$300.0	0	1	0	0	0
Queen Room - Rack Rate	\$250.0	\$260.0	\$265.0	\$270.0	0	0	1	0	0
No Purchase	\$0.0	\$0.0	\$0.0	\$0.0	0	0	0	1	0
...									
Intermediate Parameter $\hat{\lambda}$	0.50								
Intermediate Set of Parameters $\hat{\beta}$	-0.01				-0.70	0.90	0.40	-4.00	-0.75

^{a)} $D = \{T_4, T_3, T_2, T_1\}$; $P = \{T_4, T_1\}$; $\bar{P} = \{T_3, T_2\}$

The problem set up provided in Table 5 allows the analyst to go through all the steps an iteration of the EM algorithm assumes. To begin with, the expectation step requires her to compute the expected value of the indicator variables $a_{(t)}$ for all time

periods t for which a purchase is not observed. In our case, this translates into the analyst having to quantify first the probabilities $P_{NP}(t)$ of a no purchase. The MNL formulas given in equation (4) facilitate this task, and hence, the analyst can express $P_{NP}(t)$ s as:

$$P_{NP}(T_k) = e^{V_{NP,k}} / \sum_{j,k} (e^{V_{j,k}}), \quad k \in \{3, 2\}, j \in \{\dots, J, NP\},$$

which leads to $P_{NP}(T_3)$ and $P_{NP}(T_2)$ being 5.7% and 7.4%, respectively. Substituting these probabilities in equation (15), the expected values $\hat{a}(t)$ become:

$$\hat{a}(T_3) = \frac{\hat{\lambda} \cdot P_{NP}(T_3)}{\hat{\lambda} \cdot P_{NP}(T_3) + (1 - \hat{\lambda})} = 0.054 \quad \text{and} \quad \hat{a}(T_2) = \frac{\hat{\lambda} \cdot P_{NP}(T_2)}{\hat{\lambda} \cdot P_{NP}(T_2) + (1 - \hat{\lambda})} = 0.069.$$

In this context, during the current EM iteration, the values of $\hat{a}(t)$ s being closer to zero than to one suggest that the likelihoods of customers requesting hotel products during the second and the third elemental time periods are quite low.

The expected values $\hat{a}(t)$ s assist next in computing a new set of parameter estimates λ and β during the maximization step of the EM algorithm. The optimal value λ^* that maximizes the λ -component of the expected conditional log likelihood function $E(LL|\hat{\beta}, \hat{\lambda})$ can be easily determined using equation (17) and, for the particular example of this problem, it equals

$$\lambda^* = \frac{|P| + \sum_{t \in P} (\hat{a}(t))}{|P| + |\bar{P}|} = \frac{2 + (0.054 + 0.069)}{2 + 2} = 0.53.$$

While λ^* is the result of a straightforward algebraic operation, the estimation of β coefficients requires the analyst to iteratively maximize the β -component of the expected conditional log likelihood function. The maximization procedure slightly modified to account for the presence of the $\hat{a}(t)$ s follows the MNL methodology described in Sub-section 3.4.1 and, therefore, it is not revisited here again. For

convenience, let's suppose that the maximum of the β dependent expected conditional log likelihood function

$$\left[1 \cdot \ln(P_{\$200.0(T4)}) + \hat{a}(T_3) \cdot \ln(P_{NP}(T_3)) + \hat{a}(T_2) \cdot \ln(P_{NP}(T_2)) + 1 \cdot \ln(P_{\$300.0(TI)}) \right]$$

is attained at the vector of parameters β^* . If the convergence criterion

$$\left| \left[\hat{\lambda}, \hat{\beta} \right] - \left[\lambda^*, \beta^* \right] \right| < \varepsilon \text{ is true then the EM algorithm terminates and } \lambda^* \text{ and } \beta^* \text{ are its}$$

optimal solutions. If the criterion is not met, then the entire process repeats with λ^* and β^* becoming the new $\hat{\lambda}$ and $\hat{\beta}$ until a solution is reached.

3.5. Computing Preferred Capacity Control Policies

This section focuses on how service providers in several industries allocate their perishable inventories across multiple customer segments in such a way that their expected revenues are maximized. The most widely used capacity allocation technique, that is, the Expected Seat Marginal Revenue heuristic - version b (EMSR-b), which implies that customer segment i exclusively purchases products from the product class i is investigated in Sub-section 3.5.1. The limiting assumptions that guide the computation of the EMSR-b capacity controls are relaxed in Sub-section 3.5.2 which allows service providers to determine an optimal product display policy based on their customers' purchasing behavior. In this case, the focus shifts from how many products the service provider needs to reserve for the higher willingness to pay customers to how the available products need to optimally be displayed to maximize the revenues and incentivize customers to purchase.

3.5.1. Expected Marginal Seat Revenue Heuristic - Version b

The EMSR-b (Belobaba [36], [39], [67]) heuristic assumes that the service provider sells the same perishable product to multiple customer segments which accept a predetermined differential pricing scheme. While irrelevant to solving the problem, it should be noted

that how well the service provider structures its pricing scheme and, subsequently, enforces it, is directly reflected into the service provider's profit. The proposed pricing system consists of n possible product classes each priced at an appropriate p_i , $i \leq n$, level. In the hospitality industry, the n product classes may refer to rooms associated with different sale conditions and booking restrictions. For convenience, the product classes are numbered in the descending order of their prices with the first and the last of them being, therefore, priced the most and the least, respectively. Customer demand D_i for product class i is assumed to be independent over product classes and described by the cumulative distribution function $F_i(x)$. In most applications of the EMSR-b, the demands D_i are hypothesized to come from normal distributions identified by their parameters μ_i and σ_i . In addition, the demands D_i are supposed to realize in ascending order of product classes' price, that is, the cheapest products p_1 are sold first, followed by the second to the cheapest products p_2 and so on, so that the most expensive products p_n are sold last. In this scenario, there is a one-to-one correspondence between a current sale period and the product class i that is requested during this period. This correspondence is depicted in Figure 7, which provides a graphic summary of the assumptions embedded in the EMSR-b technique.

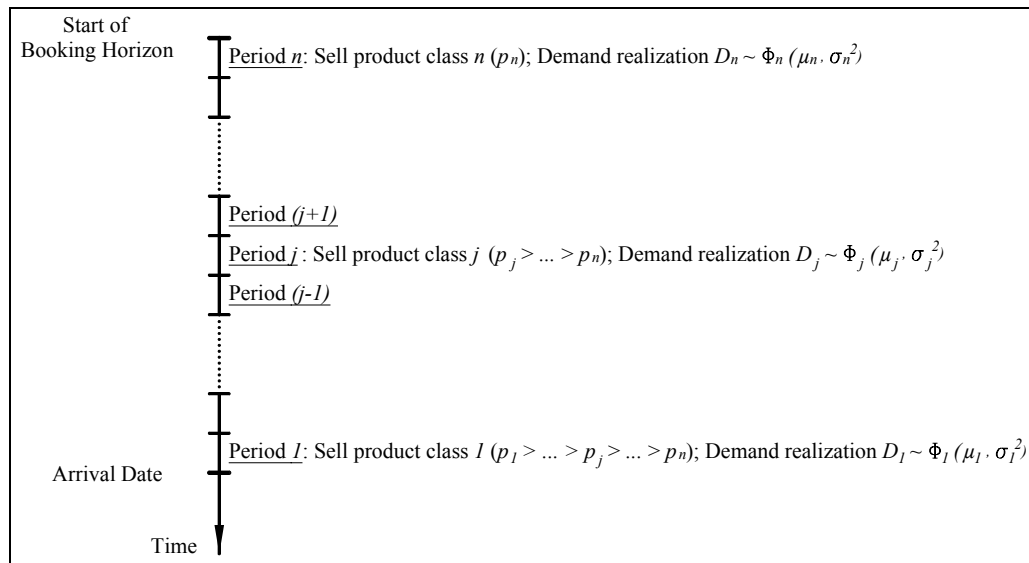


Figure 7 EMSR-b - Graphic Summary of Embedded Assumptions

During a current time period $(j+1), j < n$, the decision the service provider needs to make refers to the number of requests for product classes $j, j-1, \dots, 1$ that it should accept such that its revenues are maximized. In the revenue management parlance this is equivalent to determining the protection level θ_j for product classes $j, j-1, \dots, 1$ combined. Since working simultaneously with multiple product classes turns out to be an analytically demanding process, EMSR-b collapses all $j, j-1, \dots, 1$ product classes into an equivalent *virtual class* whose distributional parameters are computed based on the individual parameters of the corresponding product classes. Given that the demand for product classes is assumed independent, the distributional parameters of the equivalent class can be determined as:

$$\mu = \sum_{i=1}^j \mu_i, \quad \sigma = \sqrt{\sum_{i=1}^j (\sigma_i^2)}.$$

In addition, to be as representative as possible of the product classes it consists of, the virtual class is given a price that represents the weighted average price over all displaced classes, or:

$$p = \frac{\sum_{i=1}^j p_i \cdot \mu_i}{\sum_{i=1}^j \mu_i} = \frac{\sum_{i=1}^j p_i \cdot \mu_i}{\mu}. \quad (20)$$

The introduction of the virtual class transforms the multiple product class problem into an equivalent two-class capacity allocation one which is much easier to solve (Littlewood [68]). While other approaches to explaining how the optimal solution of this problem is reached exist, we present here a method that borrows from the solution algorithm of the classical newsvendor problem. In its most general form, this problem refers to the situation in which a single product, ordered at the beginning of a sale period, can be used to satisfy the uncertain demand that realizes during that period only. In such a context, the optimal product quantity Q_o to order balances the overage c_o and the underage c_u costs, that is, the unit cost of purchasing too many products and the unit cost

of purchasing too few products, respectively, and results in the following expression (see Nahmias [69] for the full proof):

$$Q_o = F^{-1} \left(\frac{c_u}{c_u + c_o} \right) \quad (21)$$

where F^{-1} is the inverse cumulative distribution function of the demand.

Adapting the newsvendor problem's concepts for our simplified problem, the intention is to determine at the beginning of each time period $(j+1)$, $j < n$, the optimal protection level θ_j in such a way that the overage and underage costs are appropriately balanced. The overage cost, or, the unit cost of accepting too many requests for product classes $j, j-1, \dots, 1$ is equal to p_{j+1} . In other words, if the protection level θ_j had been correctly set, we could have sold a $(j+1)$ class product and made an extra p_{j+1} in revenues; however, the protection level was not appropriately imposed, and this translates into a unit loss of p_{j+1} . Similarly, the underage cost, or, the unit cost of accepting too few requests for product classes $j, j-1, \dots, 1$ is equal to $(p - p_{j+1})$. In other words, if the protection level θ_j had been correctly set, we could have sold the products at p (on average) instead of p_{j+1} ; however, since the protection level was again wrongly selected, the corresponding mismatch results in an average revenue loss of $(p - p_{j+1})$ per unit. Rewriting (21) with the critical ratio $c_u / (c_u + c_o)$ as the right-hand side term of the equation, we obtain:

$$F(\theta_j) = P(D \leq \theta_j) = \frac{c_u}{c_u + c_o} = \frac{p - p_{j+1}}{p} = 1 - \frac{p_{j+1}}{p} \quad (22)$$

where $F(\theta_j)$ is the probability that the demand for the virtual class does not exceed θ_j , and the ratio $c_u / (c_u + c_o)$ is the probability of satisfying all the demand during the time period $(j+1)$ given that θ_j generic products are reserved for the product classes displaced by the virtual class at the beginning of the period. Given that the demand D for the virtual product class is typically normally distributed, equation (22) re-writes

$$F(\theta_j) = F(\theta_j; \mu, \sigma^2) = 1 - \frac{p_{j+1}}{p}$$

or, alternatively,

$$\theta_j = F^{-1}\left(\left(1 - p_{j+1}/p\right); \mu, \sigma^2\right) = \mu + \sigma \cdot \Phi^{-1}\left(1 - p_{j+1}/p\right) \quad (23)$$

where Φ^{-1} is the inverse standard normal cumulative distribution function.

The capacity allocation procedure described above compares the cost of accepting a certain request for a lower priced product of the $(j-1)$ th class to the average cost of accepting an unconfirmed request for a higher priced product of the k th class, $k \leq j$. To illustrate how the EMSR-b method is used in practice, consider the example of a hypothetical hotel which sells four types of rooms as depicted in Table 6 below. The hotel sells a maximum of 200 rooms and the unconstrained demand for each of the product classes, assumed normally distributed, is determined using traditional forecasting techniques. The room price is considered exogenous to the optimization algorithm and is set at the beginning of the booking horizon.

Table 6 EMSR-b - Input Data and Computed Protection Levels

Prod. Class	Product Class Name	Price	Unconstrained Demand		EMSR-b Protection Level θ_j	EMSR-b with Buy-Up	
			Average	Std. Dev		Prob.	θ_j
1	King Room - Rack Rate with Breakfast	\$275.0	40	6	32	0.00	34
2	King Room - Rack Rate	\$250.0	45	8	78	0.30	80
3	King Room - Advance Purchase	\$200.0	75	10	151	0.15	155
4	King Room - Ultimate Savings	\$175.0	80	12	200	0.30	200

The computation of the EMSR-b protection levels $\theta_j, j \leq 3$, starts with the sale of the rooms in the product class 4. During the corresponding time period, the future demand for product classes 3, 2, and 1 combined follows a normal distribution with parameters

$$\mu = \sum_{j=1}^3 \mu_j = 75 + 45 + 40 = 160 \text{ and } \sigma = \sqrt{\sigma_j^2} = \sqrt{6^2 + 8^2 + 10^2} = \sqrt{200}.$$

The weighted average price p associated with the virtual product class that includes classes 3, 2, and 1 is computed using equation (20) and equals \$232.8. Finally,

substituting the values of all unknown variables in equation (23) leads to the protection level θ_3 being equal to

$$\theta_3 = \mu + \sigma \cdot \Phi^{-1}(1 - p_4/p) = 160 + \sqrt{200} \cdot \Phi^{-1}(1 - 175/232.8) = 150.4 ,$$

or, if rounded up, to $\theta_3 = 151$. Applying this operation repeatedly two more times results in the protection levels θ_2 and θ_1 being evaluated at 78 and 32, respectively. Thus, for this hypothetical example, 32 rooms should be protected for the product class 1 alone, 78 rooms should be reserved for product classes 1 and 2, 151 rooms should be protected for product classes 1, 2, and 3, and the rest of 49 rooms can be offered to customers requesting product class 4. The implied relationship between the computed protection levels $\theta_j, j \leq 3$, and the product class availability is graphically shown in Figure 8 below.

Remaining Capacity	200	151	78	32	0
Accept Requests for Product Classes	4, 3, 2, 1	3, 2, 1	2, 1	1	
Product Classes Closed	-	4	4, 3	4, 3, 2	

Figure 8 EMSR-b - Relationship between Protection Levels and Product Class Availability

The EMSR-b heuristic can be modified to account for imperfect market segmentation (Belobaba [36], [39]), that is, for cases when a customer who would have normally purchased a class $(j+1)$ product actually purchases a class j , $(j-1)$, ..., or 1 product given that the product class $(j+1)$ is closed at the time the customer requests it. To put this into context, consider the example illustrated in Figure 8. Previously, a customer who hypothetically would have requested a class $(j+1)$ product when the class was closed would have been treated as a revenue loss for the service provider. In the buy-up framework, however, even if the preferred product class $(j+1)$ is closed, the customer is considered to be willing to purchase a class j , $(j-1)$, ..., or 1 product with some probability π_{j+1} , which is most of the time arbitrarily imposed. If the buy-up is considered to occur at the equivalent virtual class, that is, if π_{j+1} gives the probability that

a customer will accept the equivalent virtual product priced at p (see equation (20))
when product class $(j+1)$ is closed, then the overage cost c_o changes to $(p_{j+1} - \pi_{j+1} \cdot p)$
and equation (22) for computing the protection levels $\theta_j, j \leq n$, re-writes:

$$F(\theta_j) = \frac{c_u}{c_u + c_o} = \frac{p - p_{j+1}}{(p - p_{j+1}) + (p_{j+1} - \pi_{j+1} \cdot p)} = \frac{p - p_{j+1}}{p \cdot (1 - \pi_{j+1})} = \frac{1}{1 - \pi_{j+1}} \cdot \left(1 - \frac{p_{j+1}}{p}\right),$$

or, equivalently,

$$\theta_j = F^{-1}\left(\left(\frac{1}{1 - \pi_{j+1}}\right) \cdot \left(1 - p_{j+1}/p\right); \mu, \sigma^2\right) = \mu + \sigma \cdot \Phi^{-1}\left(\frac{1}{1 - \pi_{j+1}} \cdot \left(1 - \frac{p_{j+1}}{p}\right)\right). \quad (24)$$

The introduction of the buy-up probabilities $\pi_j, j \leq n, j > 1$ in equation (24) results in the service provider closing the lower revenue product classes more aggressively. As a result, as shown in the far right column of Table 6, the corresponding protection levels $\theta_j, j \leq 4$ exceed those computed with the simple EMSR-b heuristic. In this case, if customers are indeed willing to buy-up, the service provider is shown to gain significant incremental revenues (Belobaba and Weatherford [40]).

Given the fundamental assumptions that guide its practical implementation, the EMSR-b with buy-up represents a rather timid attempt to soundly integrate customer purchasing behavior into the traditional RM framework. Borrowing from the extensive discussion provided in Sub-section 2.4, the biggest criticism that is associated with the use of this method lies in the fact that the buy-up probabilities are not estimated from actual purchase data, but rather, they are a priori specified by some knowledgeable market analysts. With this in mind, the methodology introduced in the next sub-section relaxes many of the limiting assumptions of the EMSR-b techniques and explicitly accounts for the customer purchasing behavior when recommending an optimal capacity allocation policy.

3.5.2. Optimal Choice-Based Capacity Control Policy

This section investigates a service provider's actions that result in an optimal product display policy that maximizes the firm's expected revenues by incentivizing customers to purchase. Typically, in a revenue management framework, the revenues the service provider obtains from selling its products depend on the realized demand as much as on the capacity control policy the service provider employs to accept or deny product requests. To this end, given the stochastic character of the demand, the service provider intends to implement a robust enough control policy that guarantees it offers customers the optimal combination of products at any point in time prior to the product expiration and for any amount of the remaining capacity. The methodology that allows the service provider to compute such an optimal policy is extensively discussed in Talluri and van Ryzin [14]. For completeness, a concise summary of this methodology is provided next.

To illustrate how the optimal display policy can be determined, let's begin by assuming that the booking horizon at the service provider is discretized in T decision periods each of which is associated with at most one arrival. For convenience, the index t of a current decision period runs backwards in time, with $t=T$ representing the beginning of the booking horizon and $t=0$ identifying the expiration of the product. The probability a customer arrives during the decision period t is denoted by λ , which, for simplicity, is considered to stay the same during the entire booking horizon. The service provider is assumed to sell a maximum of n product classes, $N=\{1, \dots, j, \dots, n\}$ denoting in this context the complete set of product classes. As before, the product classes are numbered in the descending order of their prices, or, $R_1 \geq \dots \geq R_j \geq \dots \geq R_n$. In addition, for a customer facing a product subset $S \subseteq N$, her probability of purchasing product $j \in S$ is assumed to be given by a model of customer purchasing behavior of the type described in Sub-section 3.4.2.

The information provided above helps the service provider formulate a recursive solution algorithm to its problem. Suppose we are in the decision period t and the remaining capacity is x . If a product of the j th class is purchased from the product subset $S \subseteq N$ that is currently offered, then the total expected revenue that can be generated from the decision periods $\{t, t-1, \dots, 0\}$ is

$$R_j + V_{t-1}(x-1) \quad (25)$$

where $V_{t-1}(x-1)$ is the optimal total expected revenue that can be obtained from the decision periods $\{t-1, t-2, \dots, 0\}$ given the remaining capacity $x-1$. Similarly, if none of the $j \in S$ products are purchased at this stage, then the total expected revenue that can be generated from the decision periods $\{t, t-1, \dots, 0\}$ becomes equal to

$$V_{t-1}(x), \quad (26)$$

which reflects the optimal total expected revenue that can be obtained from the decision periods $\{t-1, t-2, \dots, 0\}$ given the remaining capacity x . Given that the events that result in the total expected revenues depicted in equations (25) and (26) are probabilistic in nature, the total expected revenue generated from decision periods $\{t, t-1, \dots, 0\}$ when product subset $S \subseteq N$ is displayed in the decision period t can be expressed as:

$$\sum_{j \in S} \left[\lambda \cdot P_j(S) \cdot (R_j + V_{t-1}(x-1)) \right] + (\lambda \cdot P_{NP}(S) + (1-\lambda)) \cdot V_{t-1}(x)$$

where $\lambda \cdot P_j(S)$ and $\lambda \cdot P_{NP}(S) + (1-\lambda)$ are adjustment constants that reflect the probabilities of each of the events happening. $\lambda \cdot P_j(S)$, for example, provides the probability that a customer arrives during the decision period t (λ) and, subsequently, purchases a product of the j th class ($P_j(S)$). Similarly, $\lambda \cdot P_{NP}(S) + (1-\lambda)$ gives the probability of a no purchase happening when the no purchase event is the outcome of two possible circumstances: a customer arrival (λ) followed by a no purchase ($P_{NP}(S)$), or, a no arrival at all ($1-\lambda$). Finally, iterating over all decision controls S , the optimal total

expected revenue that can be obtained from the decision periods $\{t, t-1, \dots, 0\}$ given the remaining capacity x in the decision period t writes recursively:

$$V_t(x) = \max_{S \subseteq N} \left\{ \sum_{j \in S} \left[\lambda \cdot P_j(S) \cdot (R_j + V_{t-1}(x-1)) \right] + (\lambda \cdot P_{NP}(S) + (1-\lambda)) \cdot V_{t-1}(x) \right\}. \quad (27)$$

In equation (27), if $P_{NP}(S)$ is replaced by the difference $1 - \sum_{j \in S} P_j(S)$, the optimal total expected revenue $V_t(x)$ can be successively re-written as:

$$\begin{aligned} V_t(x) &= \max_{S \subseteq N} \left\{ \sum_{j \in S} \left[\lambda \cdot P_j(S) \cdot (R_j + V_{t-1}(x-1)) \right] + \left(1 - \lambda \cdot \sum_{j \in S} P_j(S) \right) \cdot V_{t-1}(x) \right\} \\ &= \max_{S \subseteq N} \left\{ \lambda \cdot \sum_{j \in S} (P_j(S) \cdot R_j) + \lambda \cdot V_{t-1}(x-1) \cdot \sum_{j \in S} P_j(S) - \lambda \cdot V_{t-1}(x) \cdot \sum_{j \in S} P_j(S) \right\} + \\ &\quad + V_{t-1}(x) \\ &= \max_{S \subseteq N} \left\{ \lambda \cdot \left[\sum_{j \in S} (P_j(S) \cdot R_j) - (V_{t-1}(x) - V_{t-1}(x-1)) \cdot \sum_{j \in S} P_j(S) \right] \right\} + V_{t-1}(x), \end{aligned}$$

or, alternatively,

$$V_t(x) = \max_{S \subseteq N} \left\{ \lambda \cdot [R(S) - Q(S) \cdot \Delta V_{t-1}(x)] \right\} + V_{t-1}(x), \quad (28)$$

where $R(S)$ is the total expected revenue from offering set S , $Q(S)$ is the total purchase probability associated with offering set S , and $\Delta V_{t-1}(x) = V_{t-1}(x) - V_{t-1}(x-1)$ is the displacement cost associated with a unit of capacity. The boundary conditions required to solve the maximization problem in equation (28) are formally given by: (1) $V_t(0) = 0$, or, the value of zero remaining inventory units is zero at any time t , and (2) $V_0(x) = 0$, or, the value of x remaining inventory units is zero at time $t=0$. By solving the problem in equation (28), the service provider determines the optimal capacity control policy that it should follow to maximize its expected revenues. More explicitly, the service provider will be able to offer its customers at each time t during the booking horizon and for each remaining capacity x the product set S that will maximize its expected revenues.

The recursive formula depicted in equation (28) assumes that the maximization step takes place over all possible subsets $S \subseteq N$. Since a product offering consisting of n initial product classes results in $2^n - 1$ such subsets (2^n if the null set is also considered), the optimization procedure has a high computational complexity (i.e., $O(T \cdot X \cdot 2^n)$) which may limit its applicability for large n 's. Despite this apparent complexity, Talluri and van Ryzin [14] show that the optimal solution for equation (28) consists of a subset of the efficient sets only, that is, of the subsets S that “provide the most favorable trade-off between the probability of purchase $Q(S)$ and the expected revenue $R(S)$ ” (Talluri and van Ryzin [14], p. 16).

To illustrate how the efficient sets assist in finding the optimal solution for equation (28), let's consider the simplified example of a small Bed and Breakfast Inn that has a capacity of 5 rooms and accepts bookings 15 days in advance of the arrival date. For convenience, each of the 15 days is treated henceforth as a unique decision period. The vast majority of the bookings have a length of stay of 1 night and, for simplicity, customers are assumed to request service during the booking horizon at a constant rate of $\lambda=0.40$. The rooms, whose interiors are identical, can be easily upgraded or downgraded in such a way that a price differential is always warranted. Based on past experience, the management charges, on average, \$275.0, \$250.0, \$200.0, and \$175.0 per each possible room outfit per night. In this context, the decision the management needs to make refers to the set of products S that it should make available in each decision period t . In accomplishing this task, the management uses the customer-centric data that it previously collected and estimates two competing models of the type described in Sub-section 3.4.2.

The first model assumes that the utility a customer obtains from purchasing the room type j depends on the product price only. As such, the utility of product j takes on the following expression:

$$U_{nj} = \beta_P \cdot P_j + \varepsilon_{nj}, j \in \{1, \dots, 4, NP\} \quad (29)$$

where n is the n th customer, j is one of the four room outfits offered, NP refers to the *no purchase* option, P_j is the price of the j th room type, β_P is the parameter estimate associated with price, and ε_{ni} is a random component assumed error type 1 distributed. In this formulation, the utility U_{NP} of a *no purchase* is 0. For the specific problem that it tries to solve, the management obtains a price estimate of $\beta_P = -0.005$.

The second model assumes that the utility a customer obtains from purchasing the j th room outfit is influenced besides the price itself by the set of restrictions that accompany the room outfit. In this context, a set of restrictions may refer to the room being offered to the customer with a queen instead of a king bed. In addition, for each of the outfits and the *no purchase* option, an alternative specific constant $ASC_{j/NP}$ is specified. The utility of the room type j is then formally expressed as:

$$U_{nj} = ASC_j + \beta_P \cdot P_j + \beta_R \cdot R_j + \varepsilon_{nj}, j \in \{1, \dots, 4, NP\} \quad (30)$$

where R_j is a dummy variable that identifies the room outfits to which restrictions apply, β_R is the parameter estimate associated with the set of restrictions, if any, and all other entities are as explained above. For illustration purposes, let's consider that the estimation of this model results in the following parameter estimates:

$$ASC_{j/NP} = \{-0.45, -0.25, 0.00, 0.25, -3.00\}; \beta_P = -0.01; \beta_R = -0.75.$$

Discrete choice models (29) and (30), together with the computed parameter estimates, allow the management at the Bed and Breakfast Inn to identify the efficient sets $S_m, m < n$, on which the optimal solution relies. In particular, following the definition provided by Talluri and van Ryzin [14], the efficient sets are located on the efficient frontier of the scatter plots constructed with the pairs of points $(Q(S), R(S))$, $S \subset N$. In this framework, probabilities $P_j(S)$ that enter the expressions of $Q(S)$ and $R(S)$ for any subset $S \subset N$ are computed based on the substitution patterns detailed in Sub-section 3.4.1. As

shown in Figure 9 below, for the two cases analyzed, the efficient sets S_m , $m < n$, are S_I , S_5 , S_{11} , S_{15} and S_I , S_5 , S_{11} , respectively.

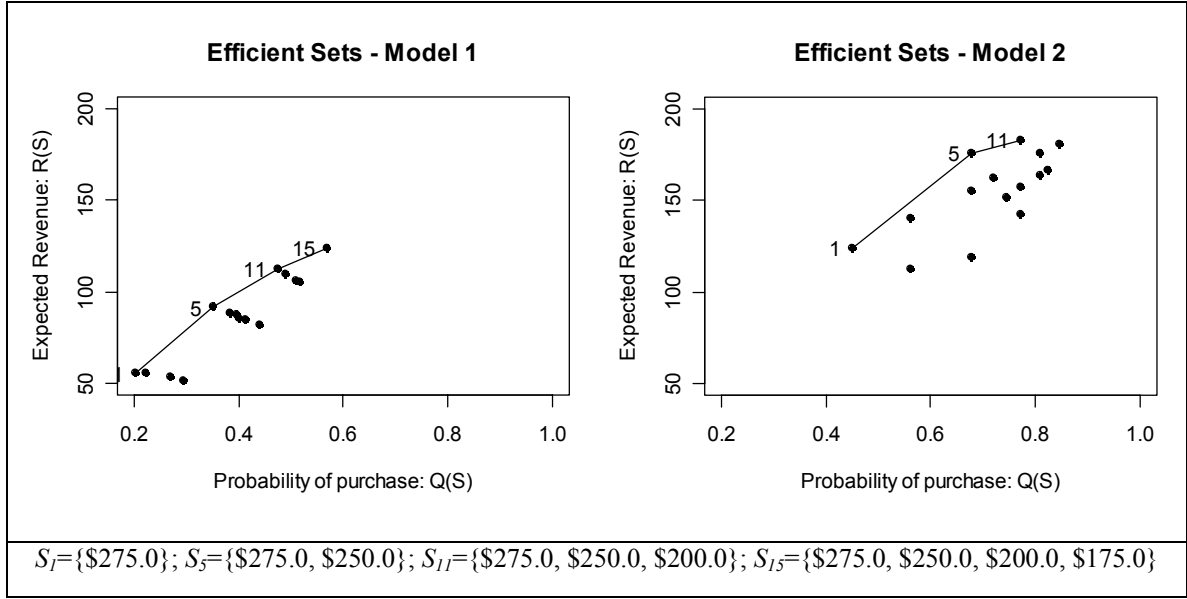


Figure 9 Efficient Frontier and Efficient Sets

The efficient sets S_m as computed above are subsequently used in the optimization step (28) which results in the optimal control policies depicted in Table 7 and Table 8. For a hypothetical stream of customer arrivals that leads to rooms being sold during the decision periods T_{13} , T_{11} , T_8 , T_5 , and T_3 , the optimal display at the Bed and Breakfast Inn is highlighted in tan in the two optimal control tables. As such, when the optimal policy is determined using the first customer purchasing behavior model, the management consistently displays the efficient set S_{15} during the entire booking horizon. In this case, continuously keeping all product classes open is justified by the high probability of a *no purchase* that accompanies this model specification. In its current formulation, given that the price always adds to the disutility of an alternative, the model specification favors unjustifiably the *no purchase* option which is associated with a constant zero utility. In contrast, when the control policy is determined based on the second customer purchasing behavior model, the management displays efficient set S_{11} , an already constrained subset, at the time the first purchase occurs. As the time passes and the capacity is sold, the management is then recommended to display efficient set S_5 which consists of the two

most expensive room outfits. In this case, the more realistic specification of the model increases the expected revenues at our hospitality establishment since the purchases occur against efficient sets for which the lower priced product classes are aggressively closed.

Table 7 Optimal Control Policy for Model 1

Decision Period															Remaining Capacity
T ₁₅	T ₁₄	T ₁₃	T ₁₂	T ₁₁	T ₁₀	T ₉	T ₈	T ₇	T ₆	T ₅	T ₄	T ₃	T ₂	T ₁	
S ₅	S ₅	S ₅	S ₅	S ₅	S ₅	S ₅	S ₅	S ₅	S ₁₁	S ₁₁	S ₁₅	S ₁₅	S ₁₅	S ₁₅	X=1
S ₅	S ₅	S ₁₁	S ₁₁	S ₁₁	S ₁₁	S ₁₁	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	X=2
S ₁₁	S ₁₁	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	X=3
S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	X=4
S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	S ₁₅	X=5

Table 8 Optimal Control Policy for Model 2

Decision Period															Remaining Capacity
T ₁₅	T ₁₄	T ₁₃	T ₁₂	T ₁₁	T ₁₀	T ₉	T ₈	T ₇	T ₆	T ₅	T ₄	T ₃	T ₂	T ₁	
S ₁	S ₁	S ₁	S ₁	S ₁	S ₁	S ₁	S ₁	S ₅	S ₅	S ₅	S ₅	S ₅	S ₁₁	S ₁₁	X=1
S ₁	S ₁	S ₅	S ₅	S ₅	S ₅	S ₅	S ₅	S ₅	S ₅	S ₅	S ₅	S ₁₁	S ₁₁	S ₁₁	X=2
S ₅	S ₅	S ₅	S ₅	S ₅	S ₅	S ₅	S ₅	S ₁₁	S ₁₁	S ₁₁	S ₁₁	S ₁₁	S ₁₁	S ₁₁	X=3
S ₅	S ₅	S ₅	S ₅	S ₅	S ₁₁	S ₁₁	S ₁₁	S ₁₁	S ₁₁	S ₁₁	S ₁₁	S ₁₁	S ₁₁	S ₁₁	X=4
S ₅	S ₅	S ₁₁	S ₁₁	S ₁₁	S ₁₁	S ₁₁	S ₁₁	S ₁₁	S ₁₁	S ₁₁	S ₁₁	S ₁₁	S ₁₁	S ₁₁	X=5

3.6. System Revenue Performance Evaluation

As highlighted in Sub-section 3.5.1, the implementation of both versions of the EMSR-b heuristic requires knowledge about the demand distribution associated with each potentially available product class. Typically, this demand is described by a normal distribution whose parameters are computed from historical data. These parameters together with the revenues product classes generate help determine the protection levels a service provider needs to employ to efficiently manage its perishable inventory. While easy to implement, EMSR-b computes the protection levels based on the assumption that the customers for the n product classes arrive in n stages, one for each product class, with

customers buying the lowest priced product class arriving first, customers purchasing the second lowest priced product class arriving second, etc.

The specific order of arrival assumption is, however, not used when the revenue performance of the EMSR-b heuristics is assessed through simulation. In particular, a mixed order of arrival together with frequent re-optimization during the booking horizon is usually employed to reasonably determine how the method performs under real operating conditions. Within such a framework, however, mainly because the product class arrivals are typically simulated as homogeneous Poisson processes that are later superimposed, each customer comes labeled with the product that she supposedly purchases. In such an environment, if the protection level for product classes that yield more revenue than the current one has not yet been reached, then the customer request is accepted and the corresponding revenue transfers to the service provider. In contrast, if the protection level has been reached, then the request for service is denied and the customer is lost, that is, the customer is rate denied. In addition, whenever the demand exceeds the capacity and the entire capacity is sold out all new coming requests are rejected and treated as house denials.

In this context, computing the revenue performance associated with a single or multiple streams of simulated arrivals is not a daunting task. In short, one only needs to sum over the product of the revenue associated with a product and the number of units of that product sold during the booking horizon. Since a denial is also associated with a preferred product, EMSR-b can also help quantify the lost revenues due to the capacity controls in place at the time of booking. As an intrinsic characteristic of determining the revenue performance associated with the use of the EMSR-b heuristic, it should be noted that the customer choices are known at the time the customer arrivals are simulated. Here, the customer does not have to choose among several time-of-booking available products, but rather, she comes knowing exactly what product class she wants and she gets it only if the product class is still open.

Within the choice-based RM framework, the optimal product display is computed based on a pre-determined customer purchasing behavior model. In this case, the order in which customers arrive and request service is irrelevant, since the choices are assumed to be the outcome of a rational behavior that aims to maximize the benefits a customer obtains from any choice situation. Similar to the approach employed in the case of the EMSR-b heuristic, if quantifying the revenue performance associated with the use of a choice-based RM system is desired, multiple arrival streams are to be simulated. Within such a framework, however, an arrival does not come with a label that identifies the product the customer is going to purchase, but rather, once it is assigned an arrival time, it will automatically be associated with a preferred product that represents the outcome of an assumed rational choice situation. This means that from all the products available to her at the time of the service request, the customer will purchase the product that will provide her with the highest utility.

To make sure that the results of such a simulation exercise are consistent with those that may be observed in reality, two sources of choice variation are supported herein. First, the Poisson processes that simulate customer arrivals assign distinct arrival times to each potential customer. This translates into the customer potentially facing different consideration sets based on when the request arrives and what the remaining capacity at that time is. Second, when faced with the same consideration sets, the choices of different customers may vary since the deterministic component of the product utility as given by the purchasing behavior model is adjusted with a stochastic component that accounts for all the factors that affect the choice but for objective reasons were left out of the model. Similar to the EMSR-b heuristic, if the No Purchase alternative maximizes the utility a customer obtains from the consideration set she is facing at the time of booking then the customer chooses not to buy and, from the service provider's perspective, she is treated as a rate denial. Furthermore, when the entire capacity is sold

out, service is denied to all new coming arrivals, irrespective of what the corresponding customers may have purchased, and these arrivals are all treated as house denials.

The above discussion provides guidelines into how the revenue performance of different capacity control policies can be independently assessed through simulation. For the purpose of this study, however, since the intention is to simultaneously determine the potential benefits associated with the use of these capacity controls, a fair simulation experiment which judiciously describes the customer arrivals and does not favor any of the control policies investigated needs to be devised.

The key assumption that accompanies the design of the proposed simulation experiment is that the demand for hotel products comes from a latent parametric distribution that is unknown to the revenue manager. Despite the lack of complete information that characterizes the environment in which sales decisions take place, the manager observes multiple realizations of the demand - treated as a multivariate random variable - and can, accordingly, make a decision with respect to how she wants to sell her limited perishable inventory. The capacity control policies she proposes influence the types of products that are being sold and, implicitly, the revenues that these policies drive. Given that any particular demand realization offers her information about requests that materialize or not in products being sold, the manager can continuously refine the control policies she employs by taking into account all marketplace intelligence that was not previously available to her. The iterative process of building demand knowledge and proposing refined capacity controls results in control policies that efficiently deal with the demand patterns implied by the unknown latent demand distribution.

In the particular case of the current study, the latent demand distribution is determined from the purchase transaction level data collected/observed at the hotel level. The large variety of products hotels offer to their customers impedes, however, the computation of this distribution at the product level. Instead, products are grouped based on price, room and service similarities into distinct product classes and the demand

distribution is determined at the resulting aggregate level. In addition, given that the booking patterns vary during the booking horizon, separate demand distributions are computed for booking intervals during which the arrival rates are assumed to stay the same. The number and the length of these intervals are determined at the hotel level based on the recommendations of the analysts that oversee the revenue activities at the studied hotels. Lastly, since the demand for hotel products vary based on the day of the week of the arrival date, separate demand distributions are computed for weekday and weekend arrival dates. For illustration purposes, a hypothetical demand distribution for a single product class and a weekend arrival date is presented in Table 9 below.

Table 9 Latent Demand Distribution - Hypothetical Example

	Product Class 1, Weekend Arrival Date						
	Booking Interval [Days prior to Arrival Date]						
	43+	(28, 42]	(21, 28]	(14, 21]	(7, 14]	(3, 7]	[0, 3]
Demand Distribution							
Average Demand	2.55	2.75	3.10	3.35	4.25	4.50	5.20
Standard Deviation	2.30	2.20	2.50	2.65	3.00	3.10	3.25

Similar to how RM activities are structured at most service providers, the RM system currently in used at the hotel chain is unable to deal with the demand that potentially arrives for products whose availabilities were previously restricted. While difficult to unconstrain the demand at the product level given the variety of products each hotel sells, unconstraining the demand at the product class level seems at least as challenging since the availabilities of products that form a product class are independent of each other. In this context, the latent demand distribution computed at the product class level is assumed to provide hospitality operators with the true, untruncated demand for hotel services. However, given that revenue managers are considered to be unaware of its intrinsic characteristics, this distribution is not directly used to recommend any of the preferred capacity control policies. Instead, it is used only to feed hypothetical RM systems with appropriate arrival streams and serves as a mean to create product class booking history. Thus, multiple arrival streams with no capacity controls in place are

first generated in an attempt to build knowledge about the product class demand and provide inputs for the EMSR-b method. At this stage, the product class demand for the remaining of the booking horizon is computed at the beginning of each booking interval.

In a subsequent step, these initial product class demands are used to determine the EMSR-b protection levels at the beginning of each booking interval. The product class requests that occur during these intervals and follow the latent demand distribution are accepted or denied based on whether the protection levels for the corresponding product classes were already reached or not. Once the appropriateness of accepting requests from a complete arrival stream is investigated, the current demand information is used to update the initial product class demands. The updating process repeats over all simulated arrival streams and, since it takes into account all demand components including the turn downs, it leads to unconstrained product class demands.

As seen above, evaluating the revenue performance of the EMSR-b heuristic requires the hotel operator to complete two independent tasks. First, the revenue manager needs to rely on the latent demand distribution to build knowledge about the product class demands. Second, after the learning phase ends, the manager has to continuously readjust the recommended protection levels to take into account the demand realized up to the current time as well as the expected updated demand for the remaining of the booking horizon. In contrast, since it makes use of the customers' purchasing behavior model previously estimated, computing the revenue performance of the choice-based RM methodology entails only the use of the arrival streams on which EMSR-b was refined. In this case, a sale is made if the requested product class is among those that identify the product set shown to the customer at the time she expresses her intention to book. Given the simplicity of the decision rules that govern a sale, the revenue performance of both capacity control techniques can be evaluated using performance measures such as: total revenue, forgone revenue, unsold capacity, number and type of denied product classes, etc. For illustrative purposes, the revenue performance evaluation

of the two competing capacity control policies is graphically depicted in Figure 10 below. In addition, a detailed description of the R (R Development Core Team [70]) script that assists with the EMSR-b revenue performance evaluation is provided in Appendix A.

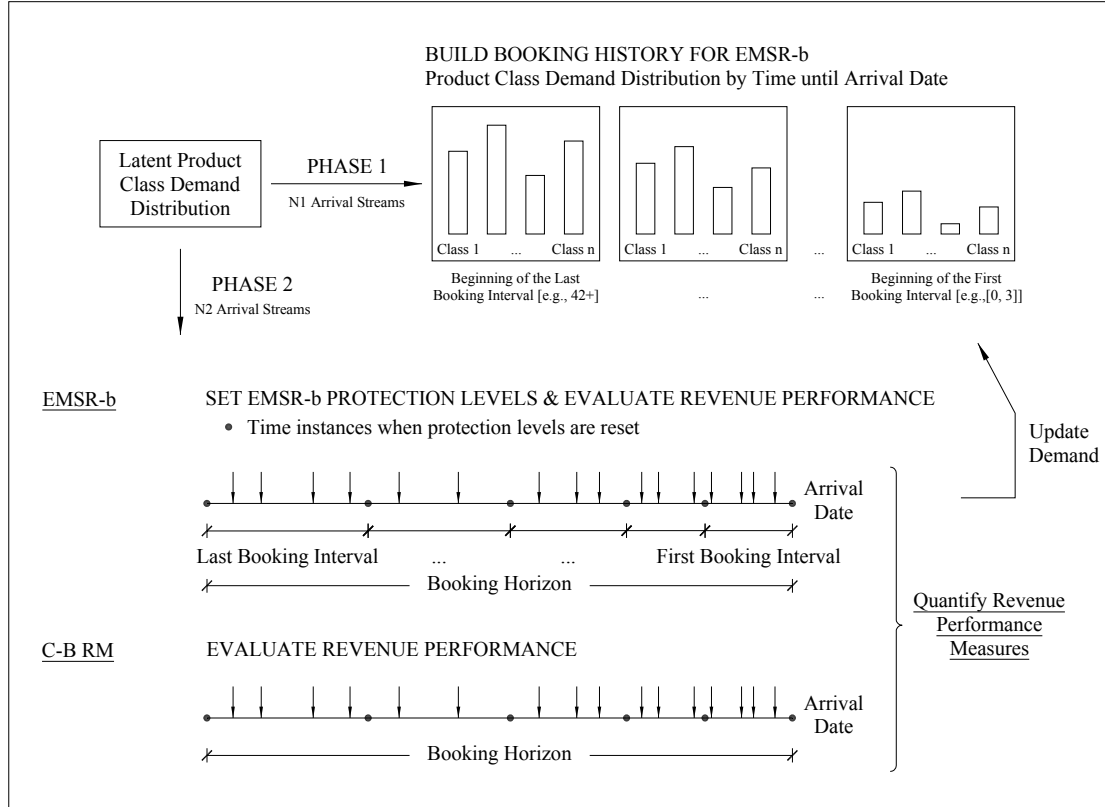


Figure 10 Revenue Performance Evaluation

3.7. Summary

This chapter presents the conceptual model that guides the development of a choice-based RM system and supports the efforts aimed to objectively quantify the revenue benefits associated with its use. Throughout the chapter, the need for customer-centric data as a prerequisite for the successful deployment of such RM systems is continuously stressed. This data, however, even if available, is shown to be incomplete in the sense that it reports the consumed transactions only while it completely disregards those that are initiated but for various reasons are not completed. To account for the incompleteness of the data, the Expectation-Maximization algorithm which replaces the

classical maximum likelihood estimation technique is employed to quantitatively describe the customers' purchasing behavior. The resulting behavioral model when integrated into the optimization routines leads to the optimal product display policy that a service provider should follow to maximize its revenue gains. In this case, the focus shifts from how much capacity should be reserved for customers willing to pay premium prices to how to optimally combine and display the available products to incentivize customers to purchase and maximize expected profits. The revenue performance of this approach is proposed to be compared to the revenue potential of the traditional RM techniques through the use of a simulated experiment intended to objectively recreate the market conditions the sponsor of this study is facing. Since the choice-based RM methodology discussed herein relies on the availability of the customer-centric data, the next chapter details the data collection plan that will enable the hospitality service provider to gather such data.

Chapter 4. DATA

It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts.

Sherlock Holmes (www.sysprog.net)

This chapter draws heavily from the work of Bodea, Ferguson and Garrow [71]. The significant contribution of all co-authors is gratefully acknowledged.

4.1. Introduction, Motivation and Purpose

Data for this study comes from a major hotel chain that wants to understand how choice-based capacity control policies may impact its long-term performance. The theoretical work of Talluri and van Ryzin [14], Zhang and Cooper [16], van Ryzin and Liu [18], and van Ryzin and Vulcano [19], [20], which acknowledges significant revenue gains associated with the choice-based control policies, constitutes the motivation for initiating this study. The challenge here is to efficiently operationalize the high level theoretical concepts that extensive simulation studies suggest to have an immediate impact on the profitability of the firm.

To date, the measurement of revenue benefits associated with choice-based RM has been based primarily on simulated data. While reported results are promising, there is a need to test these methods on real data sets. The need for a more compelling “proof of concept” is particularly important from an industry perspective, as implementation of choice-based methods will require a company to invest significant resources in

developing new RM systems. That is, choice-based systems are not simply an incremental improvement or “add-on” to existing product-based systems, but are *fundamentally different*. Consequently, successful implementation of these systems will require a company to invest significant resources in developing new data collection procedures, RM algorithms, and user support systems.

Conceptually, choice-based RM systems differ from product-based systems in that they require information about the products available to the customer at the time of booking. Collecting product availability from today’s RM systems is a daunting and time-consuming task. On the surface, the data collection methodology described in this paper may seem limited in the sense that product availability was collected for five hotel properties for 35 combinations of check-in dates and lengths of stay. However, even for this limited sample, approximately five computational hours per day over a nine week period were required to obtain accurate product availability information. In addition, numerous other scripts had to be developed to associate the correct product availability list with a specific booking and to verify data integrity. Approximately 1,100 man-hours were needed to design the data collection plan and collect and process the data. Most important, continuous interaction with more than ten experts in the hotel chain spanning revenue management, distribution channel management, system development, data management and statistics areas were required to ensure the success of this one-time data collection effort. Thus, the collection methodology and the data itself represent an important first-step towards our ability to develop choice-based RM models and benchmark their performance using more representative, real-world data. An in-depth look at the data collection process provides new insights into practical issues that need to be addressed to successfully implement choice-based RM systems in the hospitality industry.

The remainder of this chapter is organized as follows. Section 4.2 describes the steps undertaken to build a data collection plan to support the objectives of this research

study. Specifically, a data collection methodology that supplements the hotel chain's currently available purchase transaction level data with time of booking rates and room availability information is devised. Section 4.3 summarizes key findings and potential uses of the data. It also elaborates on things that interested firms should address prior to implementing choice-based RM systems.

4.2. Data Collection

In contrast to traditional RM systems, a choice-based RM system considers the availability of all products sold by the hotel of interest at the time of booking (similar to traditional RM systems, products offered by competitors are still not considered). This study developed a process to identify products available for sale at the time a booking was created. Sub-section 4.2.1 details this data collection process. Sub-section 4.2.2 describes the measurement capability of the proposed data collection plan. Sub-section 4.2.3 provides key summary statistics for the data. Sub-section 4.2.4 describes the data fields. Finally, sub-section 4.2.5 summarizes the strengths and limitations of the data.

4.2.1. Data Collection Methodology

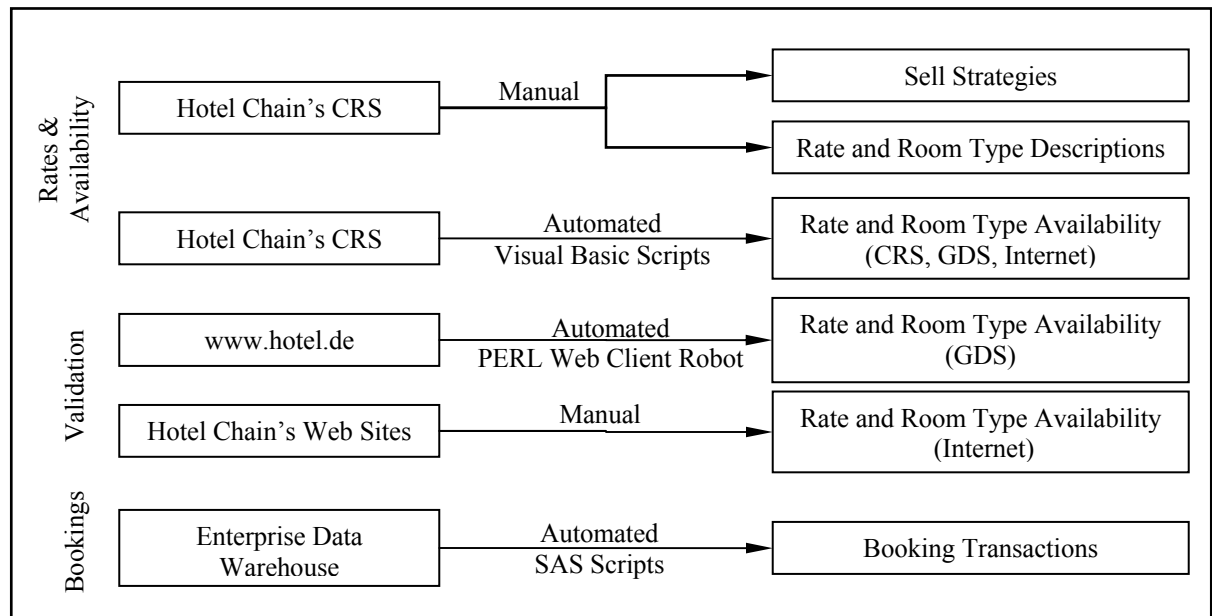
Data was collected for five hotel properties located in the continental U.S. As shown in Table 10, the properties, selected by the hotel chain sponsoring this study, were chosen to span a range of customer segments, hotel sizes, locations, and price variation (reflected in the number of hotel sell strategies, discussed later in this section). A priority was placed on collecting data from hotels that predominately serve transient business customers (vs. group and negotiated accounts customers) with short lengths of stay. By focusing on these properties, the amount of time required to collect the data was shortened, as the relevant booking horizon for such hotels is typically less than four weeks (compared to a relevant booking horizon of up to three months for leisure properties).

Table 10 Characteristics of the Study Hotels

Hotel ID	Customer Segment (Income Level)	Location	# Rooms	# Regular / Special Event Hotel Sell Strategies
1	High	Urban/Downtown	670	2 / 3
2	Medium	Suburban/Roadside	60	1 / 0
3	Medium	Suburban/Airport	160	2 / 0
4	Medium	Highway/Roadside	70	1 / 0
5	Medium to high	Urban/Downtown	260	2 / 5

As shown in Figure 11, the data collection process can be divided into three distinct phases:

1. Determining product availability lists,
2. Validating the accuracy of the product availability lists across distribution channels,
3. Obtaining booking transaction information.

**Figure 11 Data Collection Process**

To understand how product availability is determined, we first define a hotel product and explain how its availability differs by distribution channel. A hotel product is defined by a room type and rate. Room types are defined by size, view, floor height and several other characteristics (e.g., a smoking room with two queen beds with a preferred view that is wheelchair accessible). A rate contains price, restriction, and customer eligibility information. Examples of rates include:

- A price of \$150 that is available to the public that requires a 21-day advance purchase and 24-hour cancellation notice,
- A price of \$140 that is available to members of the American Automobile Association,
- A price of \$125 that is available only to employees of a large consulting firm that guarantees last room availability, that is, even if the lowest publicly-available price is \$450, employees of this firm may request and receive the pre-negotiated price of \$125.

Hotel products are sold via three distribution channels: the hotel chains' central reservation system (CRS), global distribution systems (GDSs), and the Internet. The CRS is used directly by the hotel reservation staff and the reservation personnel employed in several central reservation offices (CROs) worldwide. Off-line and on-line travel agencies use exclusively certain GDSs to get hotel and rates and room availability information. Finally, guests can directly book their hotel stays on the hotel proprietary websites. The sales process, by distribution channel, is schematically depicted in Figure 12.

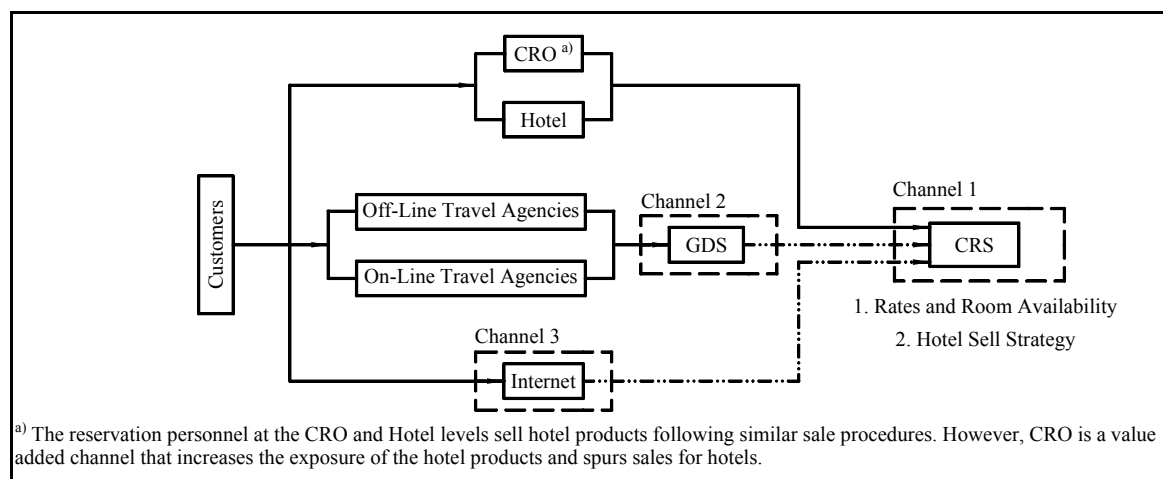


Figure 12 Selling Process by Distribution Channel

The products available for sale and the order in which these products are shown and/or verbally communicated to the customer are governed by a sell strategy. Sell

strategies vary by hotel and can change during the booking horizon. Different hotel sell strategies are commonly used for weekday vs. weekend check-in dates. Special event sell strategies may be used for holidays or special events such as the Super Bowl. Table 11 shows how selling strategies influence the number and ordering of products shown in each distribution channel. Due to confidentiality reasons, the specific rules that govern the product availability and display order cannot be made public; however, these rules were taken into account and are reflected in the recreated data sets. Bookings made via on-line travel agencies are not included in the data sets, as it was not possible to recreate available product lists at the time of booking for the corresponding distribution channel (one of the primary reasons was that it was not possible for the authors to gain access to the written contract terms between the hotel chain and the on-line travel agencies).

Table 11 Product Availability and Display Order by Distribution Channel

Distribution Channel	# Rates	# Room Types	# Products (max)	Product Order
CRO	All	All	All	Determined by hotel sell strategy
Internet	7	All	All rooms offered for the available rates	Determined by hotel sell strategy (Lowest prices appear first)
GDS (Off-line Travel Agencies)	4	3 per rate	12	GDS-specific
GDS (On-line Travel Agencies)			Difficult to recreate (terms in written contracts)	

Rate and room availability information was collected for future check-in dates occurring between Monday, March 12, 2007, and Sunday, April 15, 2007. Data for one-night stays were collected, yielding a total of 35 arrival date / length of stay combinations for each hotel. The data collection, which began on February 12, 2007, ensured a minimum booking horizon of four weeks for every hotel check-in date. Rates and room availability information was “continuously” obtained during the booking horizon by using an automated Visual Basic script (Microsoft Corporation [72]) that queried the hotel’s CRS. During the first two weeks of the data collection period (February 12 – February 25), rate and room availability was collected four times per day, or every six

hours. The frequency was increased on February 26 (or at least two weeks prior to every check-in date) to better capture the dynamics of room availability, i.e., room availabilities tend to be updated more frequently two weeks prior to check-in. With one minor exception, the frequency was increased to every three hours per day during the last seven weeks of the data collection period (February 26 - April 15) for a total of seven daily queries (the eighth query scheduled for the middle of the night was not executed as the hotel chain did not want queries against their CRS during the period when daily, system-wide updates to the RM system occur).

To determine which of the available products were shown to travel agents and/or displayed on the web, sell strategies were obtained from the hotel CRS. Because sell strategies and codes used to describe hotel products are updated infrequently, a weekly manual process was used to verify that no changes had occurred. Due to the complexities involved in determining product availability, a validation process was also designed to verify the accuracy of the recreated product lists sold via the GDS (for travel agents) and on the hotel's proprietary web sites. Specifically, an automated web client robot written in Perl (ActiveState Software Inc. [73]) was used to verify the content shown in the GDS. This script, which ran at the same time that queries were made against the CRS, showed that the content between the two sources was a perfect match. In addition, a weekly manual process was used to verify the accuracy of the content on the hotel's proprietary website and these checks also indicated a perfect match. This validation process was used to verify which products were available for sale, not the ordering of these products. However, since ordering of products is rarely (if ever) used in traditional and choice-based RM systems, the content validation was considered sufficient.

The final step of the data collection process involved obtaining booking information and associating it with the appropriate product availability list. SAS scripts (SAS Institute Inc. [1]) were used to obtain all transactional information associated with a booking from a data warehouse while STATA scripts (StataCorp. [74]) were used to

merge booking and product availability information. Only “consumed” bookings resulting in the customer staying at the hotel are included in the data set, i.e., bookings resulting in cancellations or no-shows are excluded. In order to identify potential biases in the data, all consumed bookings with check-in dates during the study period of interest are included in the recreated data sets. Descriptive statistics, detailed in sub-section 4.2.3, identify biases in the data sets that researchers should be aware of when using the data. The biases are due to the data collection design and assumptions required to merge bookings with product availability lists.

To summarize, key information contained in the recreated data sets includes:

- Transient customers (predominately business travelers) with check-in dates between March 12, 2007, and April 15, 2007 in one of five continental U.S. hotels,
- A minimum booking horizon of four weeks for each check-in date,
- Rate and room type availability information present at the time of booking for reservations made via the hotel chain’s CRS, GDSs, and proprietary web sites.

Information that is excluded from the data sets includes:

- Bookings that were cancelled or no-showed,
- Bookings made via on-line travel agencies,
- Bookings made using group, negotiated account, and other non-public rates.

4.2.2. Data Collection Plan - Measurement Capability

As with all data-driven processes, the quality of the collected data influences the quality of the results and, subsequently, the magnitude of the impact the decisions made based on these results have in real world situations. Acknowledging the importance of the data collection process, this sub-section takes a look at the measurement capabilities that the data collection plan described above shows. In particular, issues related to the precision (i.e., repeatability and reproducibility), the accuracy and the stability of the data and the

data collection process as they apply to the hotel chain's particular situation are discussed in detail.

The proposed data collection process is **repeatable** only at the time of the data collection. Since the CRS does not store any product information once it becomes obsolete, that is, once the booking date consumes, any end-user that might want to re-collect data for an arrival date that meanwhile passed will not be able to accomplish such a task. Furthermore, given that the information available at the CRS level is dynamically controlled by the RM system, even if the check-in dates of interest did not consume, the end-user might end up collecting different data for the same arrival date at different times during the data collection activity.

The data collection process is **reproducible** at the time of the data collection since the collection mechanism is fully automated. Any end-user who has access to the hotel chain's CRS and is in the possession of the data collection scripts will essentially obtain output files with similar content.

The rates and room availability information are **accurate** since the data collection scripts work directly with the CRS whose content subsequently propagates in all distribution channels. In addition, as mentioned in Sub-section 4.2.1, the accuracy of the re-created information that tentatively gets displayed at the GDS and the Internet levels is validated with the help of supplemental data collected from alternative data sources.

Finally, the data collection process is **stable** since the information the same end-user collects in repeated, shortly-distanced, consecutive trials is the same.

4.2.3. Descriptive Statistics

During the entire booking horizon a total of 17,837 Rate Category - Room Type products were booked and subsequently consumed at the five study hotels for check-in dates between March 12, 2007 and April 15, 2007. This number does not reflect the reservations that were cancelled prior to the consumption of the hotel stay and those that

were neither cancelled nor honored. Of the total of 17,837 consumed reservations, 5,172 (~29%) public rate bookings qualified for further analysis. The rest, or, 12,665 (~71%) consumed bookings were excluded because they were associated with rate types other than the public ones (e.g., 4,003 (~22%) group rates, 4,582 (~26%) negotiated accounts rates, etc.) and in most of these cases the customer did not go through a decision/selection process when she purchased her preferred product but rather accepted the product that was offered to her. The bookings that came through on-line travel agencies (i.e., 1,883 (~11%) reservations across all hotels) and were the outcome of a decision process were screened out because of the inability to collect the corresponding time of booking rate and room availability information. The consumed reservations by rate type and hotel are summarized in Table 12 below.

Table 12 Consumed Bookings by Rate Type and Hotel for Study Check-In Dates

Rate Type	Hotel 1	Hotel 2	Hotel 3	Hotel 4	Hotel 5	Total
Public	2,242	603	1,644	373	310	5,172 (29%)
Negotiated Accounts	3,015	46	507	114	900	4,582 (26%)
Group	2,356	0	746	67	834	4,002 (22%)
On-line Travel Agency	1,434	22	293	12	122	1,883 (11%)
AAA/AARP	90	284	0	116	29	519 (3%)
Other	714	33	258	297	376	1,678 (9%)
Total	9,851	988	3,448	979	2,571	17,837

Figure 13 shows several potential sources of bias in the recreated data sets. The first source of bias is due to the data collection design, which used a truncated booking horizon. Only 11.9% of all consumed bookings purchased using public rates (referred to hereafter as “bookings”) had a booking date prior to the initial collection period in the study. This is reasonable for the five hotel properties and not considered a limitation due to the fact that the majority of RM simulations use a truncated booking horizon and apply assumptions about the number of bookings present prior to the start of the simulated booking horizon. Other reasons for why bookings did not have a corresponding product list include illogical dates (e.g., the booking date is reported as occurring after the check-in date) at 3.0%, failed queries at 2.0% and forced sales at 4.7%. Forced sales occur

when a hotel sells its products directly and does not make them available for sale via the CRS, GDSs, and the Internet. Overall, these data statistics suggest the data collection design is robust.

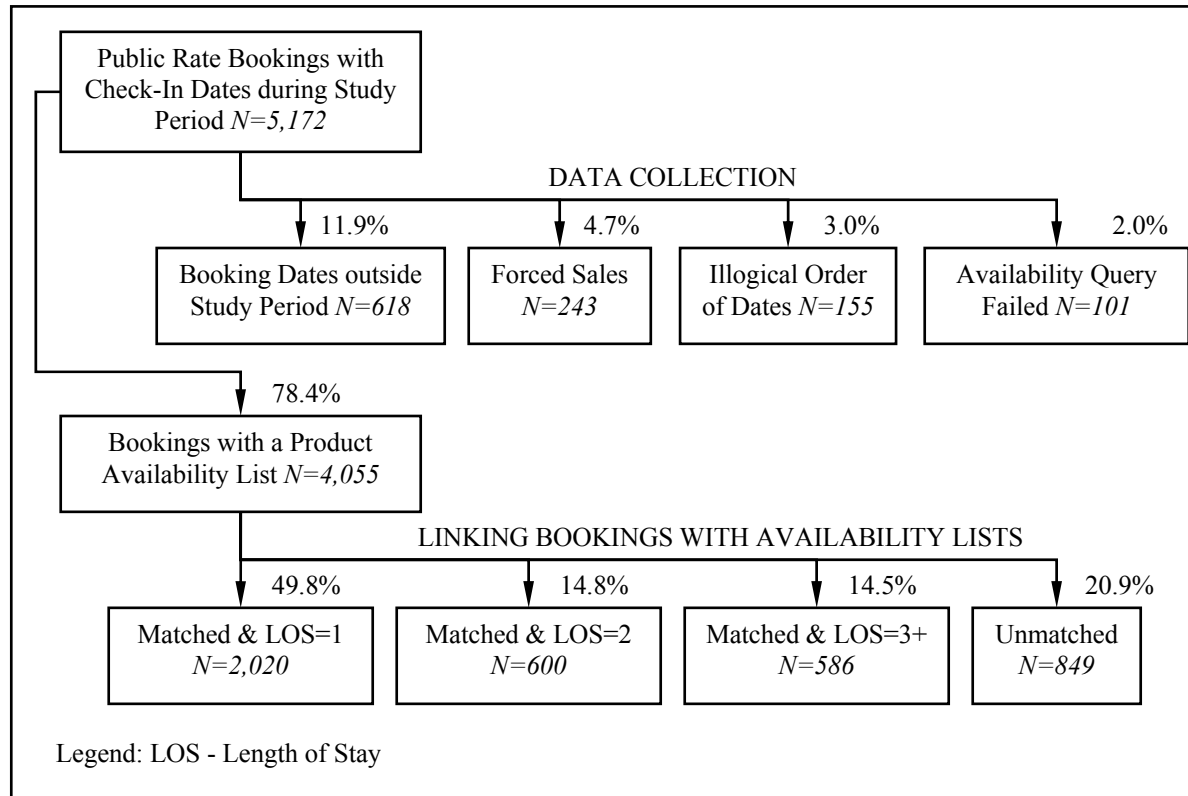


Figure 13 Potential Sources of Biases

Overall, 79.1% of bookings that had a product availability list were successfully matched. A “mismatch” occurs when a booked product can be associated with a product availability list, but does not appear in the list. While the potential sources of bias due to the data collection plan are known, many of the potential sources of bias due to the inability to successfully match bookings with a product availability list are not known. Mismatches may occur for a variety of reasons including changes that occur to the booking at the time of check-in (e.g., due to an upgrade request) or changes that occur to the product availability list during the hours between which a booking was made and

availability was recorded¹². While the exact reasons for mismatches could not be determined, it is still possible to identify potential sources of biases by comparing characteristics of bookings successfully matched with a product availability list to those from bookings that were not matched (where applicable, Chi-Square tests can be used to confirm that the observed differences were statistically significant). Based on the comparative analysis, no major differences by distribution channel were identified. The primary differences are listed below:

- Unmatched bookings are more likely to have lower prices,
- Unmatched bookings associated with hotels serving high-income travelers are more likely to have multi-night stays,
- Unmatched bookings associated with hotels serving medium-high to high income travelers are more likely to occur for bookings with shorter advance purchase times.

Additional sources of error may also be present in the data due to an underlying assumption used to match bookings with the corresponding product availability lists. Specifically, availability queries were based on a length of stay of one night. However, the consumed bookings contain both one-night and multiple-night stays. In the recreated data sets, consumed bookings with multiple-night stays were merged with product availability lists obtained from queries using one-night stays. This assumption was recommended by experts from the hotel chain sponsoring the study due to their belief that few differences in product availability are observed between lengths of stay of one or two evenings. In addition, by matching as many bookings as possible with product availability lists, richer insights may be obtained by further segments of the data (such as distribution channel). Also, this assumption provides the flexibility to other researchers who may use the data in the future to decide, based on their specific application, whether

¹² An analysis of the number of changes associated with the lowest available price offered in the distribution channel suggests that the frequency of obtaining product availability information was sufficient. A maximum of five changes to the lowest available price was observed during a booking week. The average length of time between changes to the lowest available price ranged from 5 to 35 days.

it is desirable to exclude observations with multiple lengths of stay. Finally, statistical tests indicate that the distributions of lengths of stay for matched and unmatched products are similar for Hotels 2, 3, and 4 that target medium-income customers; that is, the primary bias in the assumption is isolated to hotels serving high-income customers. In this case, proportionately more bookings with multiple night stays are unmatched. Summary statistics for the number of bookings associated with available product lists are shown in Table 13.

Table 13 Data Descriptive Statistics

Measure	Hotel 1		Hotel 2		Hotel 3		Hotel 4		Hotel 5	
	Wkday	Wknd	Wkday	Wknd	Wkday	Wknd	Wkday	Wknd	Wkday	Wknd
# (%) matched bookings	1315 (59%)		211 (35%)		1147 (70%)		288 (77%)		245 (79%)	
# matched by distribution channel										
CRO/Hotel	486	227	76	41	596	195	123	60	51	80
GDS	210	33	13	4	56	4	9	3	19	11
WEB	219	140	53	24	229	67	70	23	33	51
Number matched in which the lowest available price in the distribution channel was booked										
CRO/Hotel	177	31	44	21	133	61	61	18	23	26
GDS	139	16	9	4	18	2	2	2	8	4
WEB	103	49	34	19	160	42	42	14	25	40
Average number of days prior to the arrival date booking was made										
CRO/Hotel	10.9	15.4	5.8	4.2	5.2	6.9	2.7	7.5	3.9	5.3
GDS	10.3	12.4	7.9	7.3	9.4	6.3	7.9	17.7	6.1	4.9
WEB	13.5	17.2	6.5	5.8	7.9	9.5	5.2	8.5	6.0	6.9
Average length of stay at the hotel [days]										
CRO/Hotel	1.38	2.11	2.30	1.56	1.23	1.21	2.86	1.60	1.82	1.24
GDS	1.71	2.85	1.62	1.50	1.32	1.00	1.44	2.67	1.32	2.73
WEB	2.36	2.16	1.70	3.29	1.24	1.28	1.64	1.74	1.85	1.49
Nightly Rate: Average/Standard Deviation [\$/night]										
CRO/Hotel	401.30	334.82	158.10	147.85	114.93	103.84	112.23	105.17	201.73	117.38
	81.08	72.69	19.10	13.21	24.97	27.80	8.11	8.66	67.43	24.06
GDS	405.00	296.15	156.99	139.50	126.32	100.25	111.51	97.28	253.53	114.09
	65.61	34.13	12.85	9.00	14.62	28.81	13.62	2.89	15.62	17.5
WEB	394.62	320.85	151.20	143.63	114.55	98.04	108.42	101.95	238.91	106.02
	71.37	65.46	12.87	8.17	17.84	25.21	13.04	6.07	50.22	18.7

NOTE: Wkday - weekday check-in date; Wknd - weekend check-in date. There are a total of 25 weekday and 10 weekend check-in dates in the sample.

To summarize, the following characteristics should be kept in mind when using this data:

- All bookings with check-in dates during the study period are included,

- Bookings that were successfully matched to product lists, as well as bookings that were not successfully matched to product lists, are included in the data to allow researchers to quantify effects associated with omission bias,
- Bookings with multiple-night lengths of stay were matched with product lists obtained from queries that assumed a one-night stay.

4.2.4. Data Description

Five data sets, available in comma-separated values format, are recreated - one for each hotel. Each data set is stored in the “IDCASE-IDALT” format, one of the standard formats used to estimate discrete choice models. In this application, each “case” represents a purchased booking (defined by the Booking ID) and each “alternative” - or row in the data - represents a product available at the time of booking (defined by the Product ID). Product IDs correspond to the order in which products were displayed in the distribution channel (e.g., the first product in the availability list always has a Product ID of one). Figure 14 provides an example of some of the fields in the data sets. Note that, consistent with the IDCASE-IDALT format, characteristics describing the purchased booking and customer are the same for all rows associated with a Booking ID. In this example, Booking 44 has 22 products in its availability list. Bookings 45, 46, and 47 only have one row in the data (or a single product ID) because they could not be matched with an available product list. This is also evident by the fact that product availability information associated with these bookings is not populated, and the merge indicator is set to zero.

Booking ID	Product ID	Check-in Date	Distribution Channel	Nightly Rate	Merge Ind.	Arrival Rate	Rate Type	VIP Member
44	1	20070327	CRO/Hotel	469	1	469	Rate 2	0
44	2	20070327	CRO/Hotel	469	1	439	Rate 2	0
44	3	20070327	CRO/Hotel	469	1	529	Rate 2	0
...
44	22	20070327	CRO/Hotel	469	1	639	Rate 8	0
45	1	20070411	CRO/Hotel	399	0			0
46	1	20070322	CRO/Hotel	229	0			0
47	1	20070324	WEB	270	0			0

Purchased Booking Info
(Does not vary across Booking ID)

Product Availability Info
(Unique for each
Booking ID - Product ID)

Customer Info
(Does not vary
across Booking)

Figure 14 Data Set Format for Hotel 1, Bookings 44-47

All variables available in the data sets are described in Table 14. For confidentiality reasons, generic codes for hotel products were created. For example, if a hotel sells two distinct King Bed room types, the original room names were recoded as King Room 1 and King Room 2.

Table 14 Description of Data Set Fields

NAME	DESCRIPTION
Hotel ID	ID associated with the hotel property (described in Table 10).
Booking ID	ID associated with a booking. Begins at 1 for each hotel property.
Product ID	ID associated with a room type and rate available at time of booking. Begins at 1 for each Booking ID.
<i>ATTRIBUTES ASSOCIATED WITH THE PURCHASED PRODUCT</i>	
Booking Date	Date the booking was created (e.g., 20070403 = March 3, 2007)
Check-In Date	Check-in date (e.g., 20070307 = March 7, 2007).
Check-Out Date	Check-out date (e.g., 20070310 = March 10, 2007).
Distribution Channel	Booking distribution channel (CRO\Hotel, GDS, WEB).
Advance Purchase	Number of days prior to check-in the booking was made. There are cases in which the booking date occurs after the check-in date; these records have a value of -1.
Party Size	Number of adults and children associated with the booking.
Length of Stay	Length of stay / number of nights (e.g., 3).
Number of Rooms	Number of rooms booked (e.g., 2).
Nightly Rate	The average nightly rate the customer pays in USD (e.g., \$199.99).
Total Revenue	The total revenue in USD associated with the booking (e.g., \$2,116). It takes into account the length of stay and the number of rooms.
Purchased Rate Code	Code describing the purchased rate (e.g., Rate 1, Rate 2, etc.) Rate 1 - advance purchase; Rate 2 - rack; Rates 3 and higher are hotel-specific.
Purchased Room Type	Code describing the purchased room type (e.g., King Room 1, Queen Room 1).

Table 14 Continued

NAME	DESCRIPTION
<i>ATTRIBUTES ASSOCIATED WITH PRODUCTS AVAILABLE AT TIME OF BOOKING</i>	
Merge Indicator	Indicator variable equal to 1 if product list was found for Booking ID, 0 o.w.
Arrival Rate	Nightly rate associated with the Product ID. Missing when Merge Ind. = 0.
Rate Code	Code describing the Arrival Rate (e.g., Rate 1, Rate 2, etc.) Rate 1 - advance purchase; Rate 2 - rack; Rates 3 and higher are hotel-specific. Null when Merge Ind. = 0.
Rate Description	Description that explains the above rate codes (e.g., Rate 2 \equiv Rack Rate).
Room Type	Code describing the room type associated with the Product ID. Null when Merge Ind. = 0. (e.g., King Room 1, Queen Room 1).
<i>ATTRIBUTES ASSOCIATED WITH CUSTOMERS</i>	
Membership Status	Status in Rewards Program (0 - not a member, 1 - basic, 2 - elevated, 3 - premium).
Enrollment Date	Enrollment date in the Rewards Program. Non-members have missing values.
VIP Membership Status	Membership status of a VIP Rewards Program Member (0 – not a VIP, 1 - basic VIP, 2 - premium VIP member).
VIP Enrollment Date	Enrollment date in VIP Rewards Program. Non-members have missing values.

Notes (1): For CRO/Hotel, the Product IDs are presented in the order in which they are listed in the CRS. For GDS and WEB, Product IDs represent the order in which a travel agent or an on-line customer would see them on the computer screen. (2): In order to maintain the confidentiality of the data, Room Type and Rate fields do not carry the same meaning from one hotel to another (e.g., Rate 3 for Hotel 1 \neq Rate 3 for Hotel 2).

4.2.5. Discussion of Data Strengths and Limitations

The customer-centric data sets which combine for the first time in the hospitality industry the purchase transaction level data with the time of booking rate and room availability information are the outcome of a laborious process that entailed the utilization of resources of various types. In particular, approximately 1,100 man-hours were needed to design the data collection plan and, subsequently, collect and process the resulting raw data. Furthermore, familiarity of the investigators with three statistical environments (SAS (SAS Institute Inc. [1]), Stata (StataCorp. [74]), and R (R Development Core Team [70])) and two programming languages (Perl (ActiveState Software Inc. [73]) and Visual Basic (Microsoft Corporation [72])) helped speed the data related activities during the design and post data collection stages. In total, more than 5,000 lines of procedural programming code and more than 40 independent scripts were required to bring the

information available in the hotel chain's CRS and data warehouse in the final polished form. For the studied hotels, the data collection procedures led to 600 MB of plain text raw data most of which was the result of onerous screen scraping processes. Throughout all phases of this work, continuous interaction with several experts in various interrelated fields covering revenue management, distribution channel management, system development, data management and statistics ensured the success of the project. System-wise, a complete run that involved collecting rate and room availability information for the five hotels and all arrival dates of interest kept the CRS busy for periods of time ranging from 40 to 80 minutes depending on the time of day when the system was accessed. Given that the time of booking information could not be collected in real time, data integrity checks that contributed to the high number of required man-hours were routinely performed.

The amount of resources required to collect and process the customer-centric data together with the inherent deficiencies of such innovative methodologies, raises the question of whether investments in new data management technologies are worth being made or not. In particular, one may be concerned with the fact that the time of booking availability information is directly used to describe the public rate bookings only (i.e., ~29% of all consumed bookings across the five hotels). In addition, the inability to unambiguously associate the time of booking product lists to all consumed public rate reservations may also provide one with incentives to refute the claim that these technologies add value to the hotel chain's operations. Irrespective of their nature, these concerns are not, however, defensible. On the one hand, improved data collection and data processing procedures will be able to reduce to the minimum the number of public rate bookings for which a time of booking product list cannot be found. On the other hand, most RM systems in place today manage the availability of the public rates only and use them as reference points for setting the price and determining the availability of all other rate types. From this perspective, the effort of collecting time of booking

availability information is even more worth being made since this information can provide the hotel chain with invaluable inputs to its contract negotiating procedures.

To these concerns, several other data characteristics may limit the direct applicability of the findings of any subsequent studies that make use of these data sets. First, the product lists successfully associated with consumed reservations consist of all the products that were available to the customer at the time of booking through the distribution channel she used to secure her booking. These product lists, however, do not necessarily coincide with the consideration sets whose products' attributes were investigated by the customer prior to her making the decision to purchase. The failure to account for the evident difference between these notions may lead to biased results when choice-based RM methodologies that employ choice models other than the multinomial logit model are used (McFadden [59]).

Second, the time of booking rate and room availability information was not collected in real time but rather it was recreated, by distribution channel, from data collected from CRS at well specified time instances during the day. For high occupancy hotels for which the menu of available products changes frequently, such a procedure can easily result in mismatches between the purchased hotel products and the products reported available at the time of booking. Furthermore, such a procedure is fundamentally inefficient since the hotel chain needs to store all collected data irrespective of whether bookings are being made at the time or not. Due to operations reasons, enhancing the RM system at the hotel chain with real time availability recording capabilities was a daunting challenge that neither the investigators nor the hotel officials were willing to take.

Lastly, the frequency and the length of the forced sale periods impact the number of consumed reservations for which the time of booking product lists are found. Across all studied hotels, 243 public rate consumed reservations could not be associated with the corresponding time of booking product lists because no such information was made

available in the CRS. Overall, 85 arrival dates were forced sold at some point in time during the considered booking horizons. On average, the forced sale periods lasted 3.8 and 8.9 days for weekday and weekend check-in dates, respectively. For such arrival dates, maximum forced sale periods of 15 and 61 days, respectively, were observed in the data sets. While unavoidable due to how the data collection plan was implemented, the negative impact of forced sale periods could be mitigated if real time, point of sale availability recording enhancements were added to the capabilities of the current RM system. The complexity and the cost of such enhancements together with the uncertain outcomes of moving towards a choice-based RM system represent deterrents that future research and data management technologies need to address to make the service providers interested in RM applications of this type.

4.3. Summary

This chapter describes the data collection methodology devised to assist a major hospitality operator collect customer-centric data. Booking and availability data for five hotels over a maximum of nine week booking horizon for five weeks of customer arrivals is compiled into several data sets. These data sets are unique because they capture both the type and rate of the product purchased as well as the other available products (other room types and rates) available at the time the customers made their purchasing decisions. The data sets may serve as a “proof of concept” for the numerous choice based RM algorithms that have recently appeared in the literature. They also may be used as a universal test bed for evaluating choice-based RM algorithms versus traditional, non-choice-based, algorithms or against competing choice-based algorithms. The former is particularly important if the RM research community hopes to convince the users of RM systems to make the significant investments needed to convert over to choice-based RM systems.

In addition, this chapter provides details of the process needed to collect customer-centric data from our sponsoring hotel chain. The data collection process required significant manpower and resources from both the research group and the RM management team of the hotel chain; requiring over 1,100 man-hours along with the writing of multiple scripts written in several different programming languages and intensive utilization of the hotel chain's RM system. Even with this controlled experiment and significant level of effort, product availability information could not be obtained for all properties or arrival dates. Thus, our project also serves as a warning of the importance for an efficient, real-time product availability collection process to precede a system wide conversion to a choice-based RM system. Because our sponsoring hotel chain has a large and experienced staff of RM managers, we suspect that other firms and industries will face even more of a technical challenge in converting to choice-based RM algorithms. We hope this work provides some guidance for these firms. In conclusion, our sponsoring hotel chain is encouraged by the demonstrated ability to recreate the choice-sets at the time of each customer booking. As mentioned previously, however, they need a convincing comparison against traditional algorithms on real data before making the significant investments required to convert to a choice-based RM system. The results of such an attempt are provided in Chapter 5 that comes next.

Chapter 5. RESULTS

When it comes to getting things done, we need fewer architects and more bricklayers.

Colleen C. Barrett, President/Director/Secretary
at Southwest Airlines Company
(www.thinkexist.com)

This chapter draws heavily from the work of Bodea, Ferguson and Garrow [75]. The significant contribution of all co-authors is gratefully acknowledged.

5.1. Preface

This chapter contributes to our understanding of the benefits associated with the use of choice-based RM techniques. In this context, the customer-centric data described in Chapter 4 is used to illustrate the difficulties that may arise when a choice-based RM implementation is attempted. In particular, it has emerged that the revenue performance of a choice-based RM system intrinsically depends on how customer purchasing behavior models are specified. While improved performance is expected when refined behavioral models are employed, inconsistent earnings have invariably been experienced when such models have been used. The inability to relate to a preferred model specification constituted the motivation for approaching measuring the choice-based benefits differently. In the new framework, a business environment in which customers always purchase on price is investigated. Since the rules that lead to a sale are pre-specified, this environment allows the estimation of behavioral models that appropriately describe how customers make purchasing decisions. The corresponding choice-based RM policies

report incremental revenue gains that range between 1.0% and 14.0%. These revenue gains are shown to be stable with respect to how the behavioral inputs are computed and consistent in both, constrained and unconstrained capacity regimes. The methodology that leads to these findings is detailed in Sections 5.2 - 5.7. In this section, a brief description of the problems that may be associated with the use of customer-centric data collected from uncontrolled business environments prefaces the discussion of the lowest fare available model.

The hotel properties for which customer-centric data was collected offer their customers a myriad of products. To facilitate the computation of appropriate control policies, these products are grouped into several product classes or buckets based on product commonalities such as price, rate codes, etc. In this context, the purchase transaction level data collected at the hotel level helps compute the parameters of the underlying product class demand distributions. Since the revenue managers are typically unaware of how the demand is distributed, multiple realizations of the product class demand are simulated in an attempt to build product class booking history. The booking history, continuously updated with the unconstrained product class demand that identifies all newly simulated arrival streams, provides the inputs required for the computation of the EMSR-b control policies. The corresponding protection levels together with the evaluation product class arrival streams are used to assess the revenue performance of EMSR-b static and dynamic across a wide range of capacity levels. The reason why multiple capacity levels are investigated relates to the hypothesis that proportionally more revenue gains are to be observed when the capacity is constrained and a superior control policy is employed.

The purchase transaction level data together with the time of booking availability information supports the estimation of various customer purchasing behavior models. The basic model specification of Talluri and van Ryzin [14], or, *TvR formulation*, is successively refined to account for the presence of a complete set of alternative specific

constants - *Full Formulation*, an incomplete set of alternative specific constants - *Restrictive Formulation*, or, alternatively, a completely a priori model specification - *Business Rules*. Whenever appropriate, variables that reflect product class restrictions (e.g., advance purchase, etc.) are added to the refined model specifications. Due to the incompleteness of the transaction level data, the EM algorithm is used to compute the parameter estimates of the behavioral models. These models, integrated into the dynamic programming optimization routines, assist in computing the optimal display policies the hotel operator needs to follow to maximize its expected revenues. The performance of the recommended choice-based control policies is assessed using the same arrival streams that helped quantify the revenue performance of EMSR-b static and dynamic. In most cases, given the inferior performance of EMSR-b static, the revenue performance of the choice-based methods is reported relative to EMSR-b dynamic. In addition, whenever opportune, the performance of competing choice-based specifications is depicted in the same charts to favor a direct comparison. All steps of the methodological approach to comparing the revenue performance of competing RM techniques are summarized in Table 15.

Consistent across all levels of capacity investigated, the EMSR-b dynamic statistically outperforms its static counterpart (see Figure 15). In capacity constrained regimes, most choice-based formulations show revenue improvements in the range of 1.0% - 2.0% over EMSR-b dynamic. These revenue gains, however, are not always statistically significant and, more importantly, tend to be choice formulation dependent. In capacity unconstrained regimes, due to additional constraints imposed on accepting low fare product requests, the choice-based RM formulations statistically underperform the EMSR-b dynamic. In this case, with EMSR-b dynamic as the baseline, revenue losses of up to 5.5% are frequently observed. Figure 16 illustrates both of these cases.

The revenue assessment process reveals that the performance of any choice-based RM formulation depends on the importance given to the *No Purchase* alternative by the

customer purchasing behavior model. Extreme formulations (e.g., *Full Formulation*) that lead to customers' choices not being influenced at all by the presence of the *No Purchase* alternative perform the worst across all levels of capacity since the corresponding control policies recommend closing most product classes early during the booking horizon. In such cases, there is simply not enough high fare product class demand to compensate for the denied low and medium fare requests. At the other end of the spectrum, formulations that suggest customers are unlikely to buy any of the product classes offered (e.g., *Restrictive Formulation*) tend to perform below EMSR-b dynamic in capacity constrained regimes and close to EMSR-b dynamic in capacity unconstrained situations, respectively (see Figure 17 for an illustrative example). Optimal choice-based control policies that allocate the capacity on a first-come, first-served basis are responsible in these cases for the revenue losses observed in the capacity constrained regimes. In between these extremes, formulations that associate a reasonable probability to the *No Purchase* alternative (e.g., *TvR Formulation*) typically report revenue gains in the capacity constrained regimes and revenue losses for their unconstrained equivalents. Interestingly enough, as long as the *No Purchase* probability stays within reasonable bounds, how the total purchasing probability distributes among the available product classes seems to have a trivial impact on the revenue performance of competing choice-based RM formulations. This observation, supported by the results obtained from employing two a priori choice-based formulations, suggests that the scientific, highly controlled approach to developing and implementing a choice-based RM system works as fine as any other less controlled ones as long as some sensible *business rules* apply (see Figure 18). The inability to relate to a preferred model specification constitutes the motivation for shifting to business environments in which the behavior of the customers is well understood and can be appropriately modeled. The important aspects of computing optimal choice-based policies for such environments are detailed in the next sections.

Table 15 Revenue Performance Assessment - Methodology and Examples

1. Compute latent demand distributions by product class - the product classes are defined based on product commonalities such as price, rate category, etc.			
2. Build booking history by product class:			
▪ $N_1 = 25$ independent arrival streams.			
3. Implement EMSR-b method (static & dynamic).			
4. Compute revenue performance of EMSR-b static and dynamic as a function of the available capacity (Figure 15):			
▪ Product class demand fixed (i.e., Hotel 1: weekday arrival dates - ~76 units/day; weekend arrival dates - ~87 units/day),			
▪ $N_2 = 100$ independent evaluation arrival streams,			
▪ Unconstrained demand,			
▪ Available capacity varies from 50 to 120 units in steps of size 10.			
5. Estimate choice model using EM algorithm (Hotel 1)			
<u><i>TvR Formulation</i></u>	<u><i>Full Formulation</i></u>	<u><i>Restrictive Formulation</i></u>	<u><i>Business Rules</i></u>
Utility formulation: $U_j = \beta_{\text{Price}} \cdot \text{Price}_j$ $j \in \{\text{No Purchase}, 1, \dots, 9\}$ $\text{Price}_{\text{No Purchase}} = 0$	Utility Formulation: $U_j = \text{ASC}_j + \beta_{\text{Price}} \cdot \text{Price}_j + \beta_R \cdot R_j$ $j \in \{\text{No Purchase}, 1, \dots, 9\}$ $\text{Price}_{\text{No Purchase}} = 0$	Utility Formulation: $U_j = \text{ASC}_j + \beta_{\text{Price}} \cdot \text{Price}_j + \beta_R \cdot R_j$ $j \in \{\text{No Purchase}, 1, \dots, 9\}$ $\text{Price}_{\text{No Purchase}} = 0 ; \text{ASC}_{\text{No Purchase}} = 0$	Utility Formulation (\equiv Full Formulation): $U_j = \text{ASC}_j + \beta_{\text{Price}} \cdot \text{Price}_j + \beta_R \cdot R_j$ $j \in \{\text{No Purchase}, 1, \dots, 9\}$ $\text{Price}_{\text{No Purchase}} = 0$ $\text{ASC}_{\text{No Purchase}}$ a priori set to meet certain goals
Weekday arrival dates: $\beta_{\text{Price}} = -0.0012$	Weekday arrival dates: $\text{ASC}_{10 \times 1} = \begin{Bmatrix} -8.38, -1.44, 0.00, -0.97 \\ -1.71, 0.00, 0.67, -1.57 \\ -1.10, 0.00 \end{Bmatrix}$ $\beta_{\text{Price}} = -0.0145$ $R_j = 1$ if $j \in \{1, 2, 3\}$; 0 otherwise $\beta_R = -0.33$	Weekday arrival dates: $\text{ASC}_{10 \times 1} = \begin{Bmatrix} 0.00, -1.21, 0.00, -1.80 \\ -1.45, 0.00, 0.23, -1.66 \\ -1.03, 0.00 \end{Bmatrix}$ $\beta_{\text{Price}} = -0.0041$ $R_j = 1$ if $j \in \{1, 2, 3\}$; 0 otherwise $\beta_R = -0.42$	Weekday arrival dates: Business rule 1 $\text{ASC}_{\text{No Purchase}} = -5.88$ Business rule 2 $\text{ASC}_{\text{No Purchase}} = -5.50$
<u>Note:</u> <i>No Purchase</i> the most preferred alternative.	<u>Note:</u> <i>No Purchase</i> the least preferred alternative.	<u>Note:</u> <i>No Purchase</i> among the most preferred alternatives.	<u>Note:</u> <i>No purchase</i> carries the corresponding probabilities from TvR specification.
6. Implement Dynamic Programming (DP) method using all choice model formulations.			
7. Compute choice-based revenue performance using the same $N_2 = 100$ independent evaluation arrival streams (TvR Specification - Figure 16; Full Specification - Not Reproduced/Inapplicable Results; Restrictive Specification - Figure 17; Business Rules - Figure 18).			

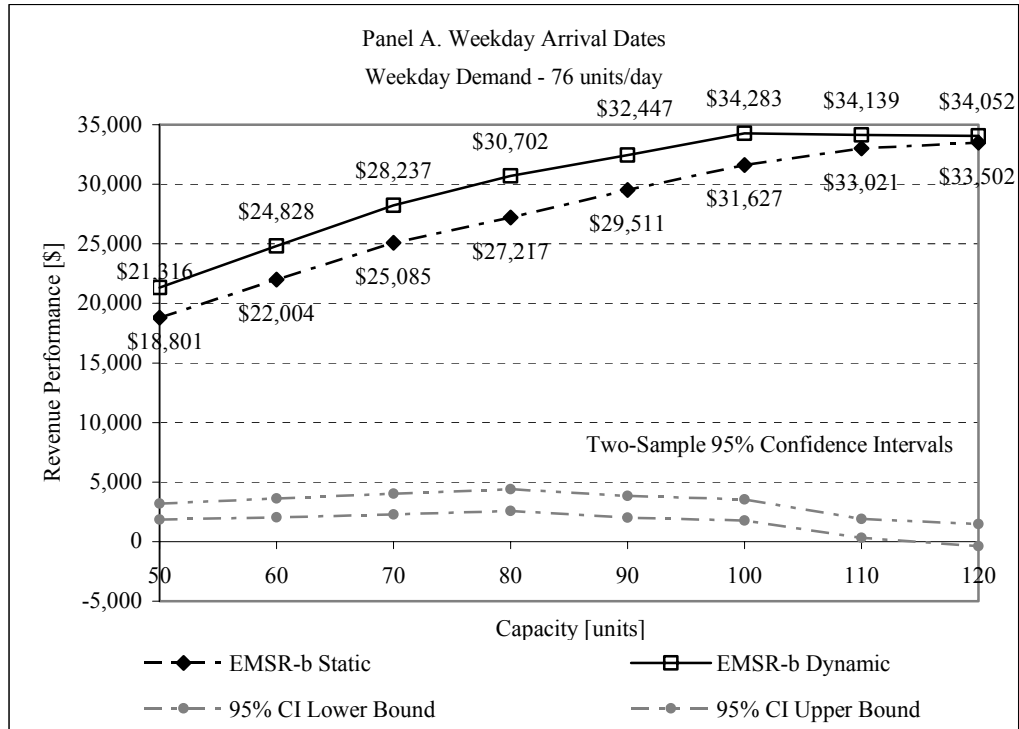


Figure 15 Revenue Performance Comparison: EMSR-b Static vs. EMSR-b Dynamic

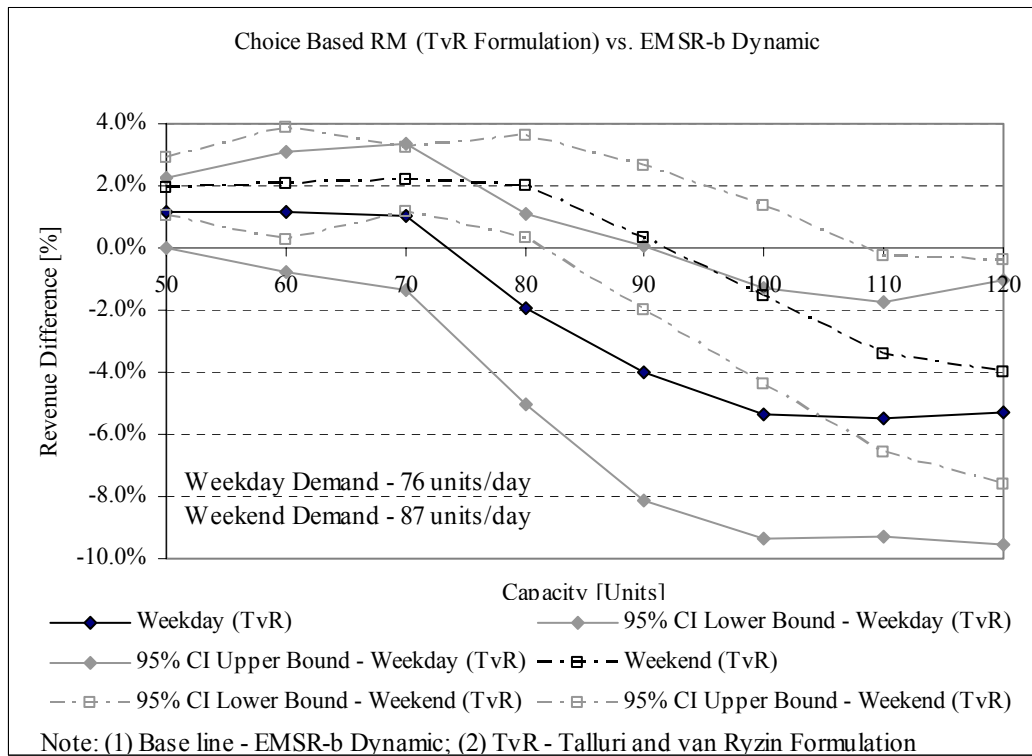


Figure 16 Revenue Differences between Choice-based RM and EMSR-b Dynamic: Two-Sample 95 % Confidence Intervals

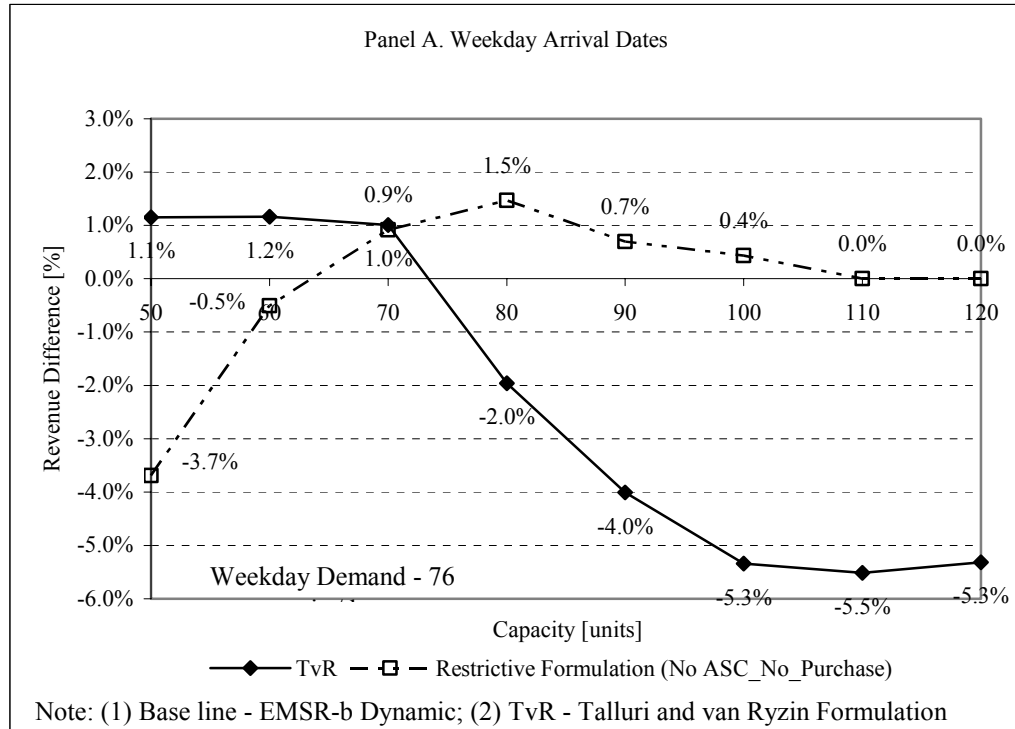


Figure 17 Revenue Performance Comparison: Competing Customer Purchasing Model Specification

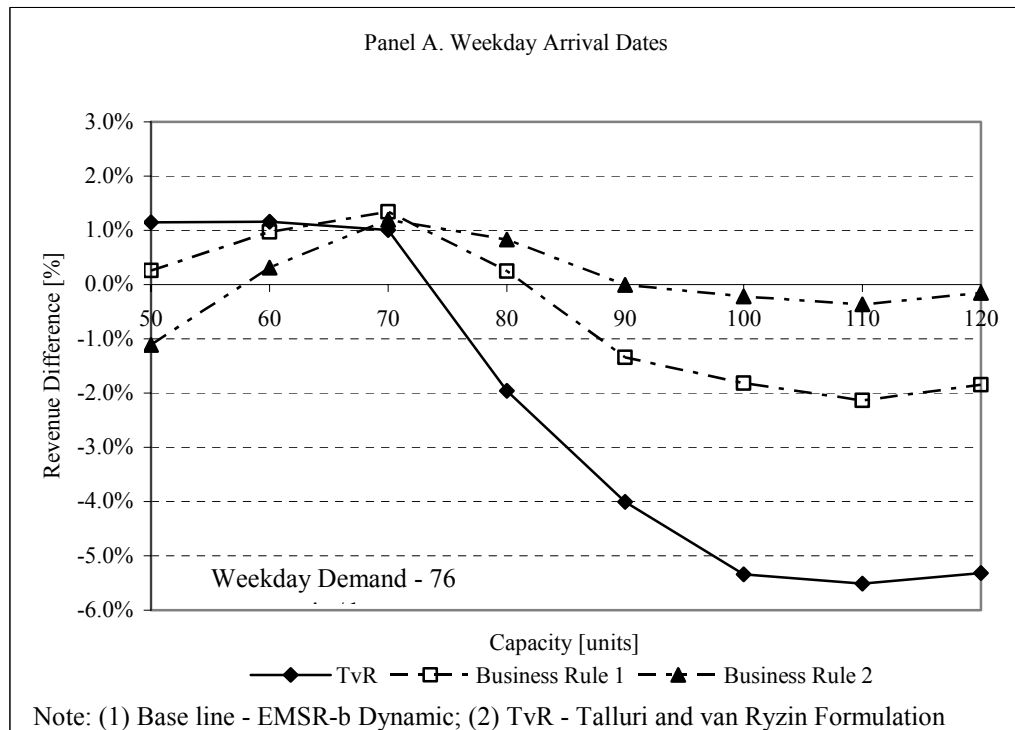


Figure 18 Revenue Performance Comparison: Competing Imposed Business Rules

5.2. Introduction and Motivation

Service providers such as airlines, hotels and car rental agencies have successfully used revenue management (RM) to control the availability of their perishable inventories. Recently, however, the business environment in which these firms have operated has seen dramatic changes. The price transparency enabled by the Internet and the discounted fares offered by on line fare consolidators have impacted the yield performance of the current RM systems and rendered them obsolete. As a result, many of the original modeling assumptions incorporated in these systems have been openly challenged and considered unrepresentative of the fast changing new business conditions. In particular, a major criticism refers to the demand for a product class being considered independent of the other product classes available to the customer at the time of booking. This assumption, which, where applicable, provides for a convenient and analytically tractable model of demand, disregards the purchasing behavior of customers who tend to consider the paid price as the only incentive to purchase. In this context, the experts' opinion (Boyd and Kallesen [5], Bakos [76], Doubleclick Performics [77], Howlett [78], Jones [79], Starkov and Price [80]) that increasingly more customers purchase on price rather than product characteristics is supported by the empirical findings of Bodea, et al. [71] who report that approximately 65% of the guests who booked on line hotel services with five US-based properties purchased the lowest fares available.

In business environments that allow customers to purchase on price the use of traditional forecasting models may contribute to the dilution of firms' revenues. Since customers cannot be segmented based on their true willingness to pay but rather on the fare they end up purchasing, the corresponding forecasts may be biased and may overestimate the demand for low fare products. As a result, more of these products are made available to potential customers and, if the same actions are consistently taken during each of the forecasting cycles, firms are shown to experience substantial revenue

losses (Boyd and Kallesen [5], Boyd, et al. [81], Cooper, et al. [82], Kuhlmann [83]). In addition, even if appropriate forecasting models are employed, revenues are still dependent on the capacity control policies RM systems recommend. To date, however, no research studies have thoroughly looked into the impact that the type of the employed RM controls has had on the revenue performance of firms that have operated in such price sensitive environments. The current research intends to fill this gap. From our experience of running this study it appears that control policies that make use of the choice-based RM concepts outperform those that rely on traditional RM methods. The primary objective of this study is to assess how competing RM techniques perform in a hypothetical setting that describes a perfect lowest fare available environment. In this context, revenues associated with the expected marginal seat revenue heuristic - version b (EMSR-b) (Belobaba [36], [39], [67]) , the choice-based RM methodology (Talluri and van Ryzin [14], [35]) , and an extension of the latter are compared in an attempt to determine whether or not one of these techniques consistently outperforms the others. For each of the methods considered, customers are assumed to fundamentally be indifferent between what various product classes represent (Boyd, et al. [81]). Thus, customers always purchase the lowest priced product available if the corresponding price is below customers' assumed willingness to pay. As implemented in this framework, offered product classes represent different prices for essentially the same product.

The remainder of this study is organized as follows. Section 5.3 provides an overview of the lowest fare available model and formalizes the mathematical concepts it employs. Section 5.4 explores the capacity control techniques investigated in this study and provides support information that facilitates their operationalization. Section 5.5 details the experimental setup. Section 5.6 describes how the recommended capacity control policies are computed and illustrates, through relevant examples, how the corresponding controls help manage the available capacity. Section 5.7 explains how the revenue performance is measured and provides insights into why the choice-based

control policies outperform the traditional RM techniques. Finally, Section 5.8 offers the conclusions of this study. With one notable exception, all processes described in Sections 5.5 - 5.7 are performed using customized scripts written in R for Windows OP (R Development Core Team [70]). The algorithms used to estimate the purchasing behavior models through the expectation-maximization technique are developed under Windows OP and coded in GAUSS (Aptech Systems Inc. [84]).

5.3. Lowest Fare Available Model

In a business environment in which customers purchase exclusively on price, the service provider offers a set of products from which customers buy the lowest priced product available if the corresponding price falls below customers' willingness to pay for service. In such a context, if j is the lowest priced product available, the customer will purchase product j with probability P_j and not purchase at all with probability $P_{NP} = (1 - P_j)$. In addition, all other products in the offered set have a zero probability of purchase. Due to the structure of the model, probabilities P_j are monotonically decreasing with the price of the lowest priced product available.

In various RM settings, the choice probabilities P_j may or may not influence the recommended capacity control policies. The traditional RM techniques, for example, completely ignore them since they make use of the demand at the product class level to suggest the recommended controls. In these instances, the product demand distributions need to be computed to account for the underlying buy-down behavior but the adjustment procedures do not require the use of P_j 's. In contrast, when customer purchasing behavior is explicitly taken into account when recommending the control policies, choice probabilities P_j are essential inputs for all subsequent computational procedures. As a result, they need to be estimated before any attempt is made to compute the control policies.

Section 5.4 that comes next discusses the competing capacity control techniques that this study investigates and, where appropriate, it provides details on how the choice probabilities P_j are estimated.

5.4. Capacity Control Techniques

5.4.1. EMSR-b Heuristic

The EMSR-b heuristic assumes that the service provider sells the same perishable product to multiple customer segments which accept a predetermined differential pricing schema. The proposed pricing system consists of n possible product classes each priced at an appropriate $p_j, j \leq n$, level. For convenience, the product classes are numbered in the descending order of their prices with the first and the last of them being, therefore, priced the most and the least, respectively. Customer demand D_j for product class j is assumed to be independent over product classes and follow a normal distribution described by parameters μ_j and σ_j . In addition, the demand distributions D_j are supposed to realize in ascending order of product classes' price. In this scenario, there is a one-to-one correspondence between a current sale period and the product class j that is requested during this period. For the purpose of this discussion, the demand distributions D_j are considered to be already adjusted to account for the customers' buy down behavior. One of the possible adjustment processes is described in Section 5.6.

During a current time period $(j+1), j < n$, the decision the service provider needs to make refers to the number of requests for product classes $j, j-1, \dots, 1$ that it should accept such that its revenues are maximized. In the RM parlance this is equivalent to determining the protection level θ_j for product classes $j, j-1, \dots, 1$ combined. Since working simultaneously with multiple product classes turns out to be an analytically involved process, EMSR-b collapses all $j, j-1, \dots, 1$ product classes into an equivalent virtual class whose distributional parameters are computed based on the individual

parameters of the corresponding product classes. Given that the demand for product classes is assumed independent, the distributional parameters of the equivalent class can be expressed as $\mu = \sum_{i=1}^j \mu_i$, $\sigma = \sqrt{\sum_{i=1}^j (\sigma_i^2)}$. In addition, to be representative of the product classes it consists of, the virtual class is given a price that represents the weighted average price over all displaced classes, or: $p = \sum_{i=1}^j p_i \cdot \mu_i / \sum_{i=1}^j \mu_i = \sum_{i=1}^j p_i \cdot \mu_i / \mu$. The introduction of the virtual class transforms the multiple product class problem into an equivalent two-class capacity allocation one which is much easier to solve (Littlewood [68]). By balancing the cost of accepting a certain request for a lower priced product of the $(j+1)$ th class with the average cost of accepting an unconfirmed request for a higher priced product of the k th class, $k \leq j$, the protection levels θ_j can be computed as $\theta_j = \mu + \sigma \cdot \Phi^{-1}(1 - p_{j+1}/p)$, where Φ^{-1} is the inverse standard normal cumulative distribution function. During the actual selling season, following the recommended protection levels, a request for product class $(j+1)$ will be accepted iff the remaining capacity at the time exceeds the protection level θ_j .

5.4.2. Choice-Based Capacity Control Technique

This section investigates a service provider's actions that result in an optimal product display policy that maximizes the firm's expected revenues by incentivizing customers to purchase. The methodology that allows the service provider to compute such an optimal policy is extensively discussed in Talluri and van Ryzin ([14], [35]). For completeness, a concise summary of this methodology is provided next.

In this framework, the booking horizon at the service provider is discretized in T decision periods each of which is associated with at most one arrival. For convenience, the index t of a current decision period runs backwards in time, with $t=T$ representing the beginning of the booking horizon and $t=0$ identifying the expiration of the product. The probability a customer arrives during the decision period t is denoted by λ , which, for

simplicity, is considered to stay the same during the entire booking horizon. The service provider is assumed to sell a maximum of n product classes, $N=\{1, \dots, j, \dots, n\}$ denoting in this context the complete set of product classes. As before, the product classes are numbered in the descending order of their prices, or, if expressed in terms of revenues, $R_1 \geq \dots \geq R_j \geq \dots \geq R_n$.

To formulate a recursive solution algorithm to the service provider's problem, let's consider we are in the decision period t and the remaining capacity is x . If a product of the j th class is purchased from the product subset $S \subseteq N$ that is currently offered, then the total expected revenue that can be generated from the decision periods $\{t, t-1, \dots, 0\}$ is $R_j + V_{t-1}(x-1)$, where $V_{t-1}(x-1)$ is the optimal total expected revenue that can be obtained from the decision periods $\{t-1, t-2, \dots, 0\}$ given the remaining capacity $(x-1)$. Similarly, if none of the $j \in S$ products are purchased at this stage, then the total expected revenue that can be generated from the decision periods $\{t, t-1, \dots, 0\}$ becomes equal to $V_{t-1}(x)$, which reflects the optimal total expected revenue that can be obtained from the decision periods $\{t-1, t-2, \dots, 0\}$ given the remaining capacity x . Given that the events that result in the above total expected revenues are probabilistic in nature, the total expected revenue generated from decision periods $\{t, t-1, \dots, 0\}$ when product subset $S \subseteq N$ is displayed in the decision period t can be expressed as:

$$\sum_{j \in S} \left[\lambda \cdot P_j(S) \cdot (R_j + V_{t-1}(x-1)) \right] + (\lambda \cdot P_{NP}(S) + (1-\lambda)) \cdot V_{t-1}(x),$$

where $\lambda \cdot P_j(S)$ and $\lambda \cdot P_{NP}(S) + (1-\lambda)$ are adjustment constants that reflect the probabilities of each of the events happening. $\lambda \cdot P_j(S)$, for example, provides the probability that a customer arrives during the decision period t (λ) and, subsequently, purchases a product of the j th class ($P_j(S)$). Similarly, $\lambda \cdot P_{NP}(S) + (1-\lambda)$ gives the probability of a no purchase happening when the no purchase event is the outcome of two possible circumstances: a customer arrival (λ) followed by a no purchase ($P_{NP}(S)$), or, a

no arrival at all $(1-\lambda)$. Finally, iterating over all decision controls S , the optimal total expected revenue that can be obtained from the decision periods $\{t, t-1, \dots, 0\}$ given the remaining capacity x in the decision period t writes recursively:

$$V_t(x) = \max_{S \subseteq N} \left\{ \sum_{j \in S} \left[\lambda \cdot P_j(S) \cdot (R_j + V_{t-1}(x-1)) \right] + (\lambda \cdot P_{NP}(S) + (1-\lambda)) \cdot V_{t-1}(x) \right\} \quad (31)$$

or, alternatively:

$$V_t(x) = \max_{S \subseteq N} \left\{ \lambda \cdot [R(S) - Q(S) \cdot \Delta V_{t-1}(x)] \right\} + V_{t-1}(x) \quad (32)$$

where $R(S)$ is the total expected revenue from offering set S , $Q(S)$ is the total purchase probability associated with offering set S , and $\Delta V_{t-1}(x) = V_{t-1}(x) - V_{t-1}(x-1)$ is the displacement cost associated with a unit of capacity. The boundary conditions required to solve this maximization problem are formally given by: (1) $V_t(0) = 0$, or, the value of zero remaining inventory units is zero at any time t , and (2) $V_0(x) = 0$, or, the value of x remaining inventory units is zero at time $t=0$. By solving the maximization problem depicted in Equations (31) and (32), the service provider will be able to offer its customers at each time t during the booking horizon and for each remaining capacity x the product set S that will maximize its expected revenues and incentivize customers to purchase.

In Equation (31), choice probabilities $P_j(S)$ are assumed to be known at the time the maximization process occurs. Within the framework built around the lowest fare available model, two approaches are proposed to compute these probabilities. The first requires the estimation of a multinomial logit (MNL) model, the most common of all discrete choice models, using available transaction level data. Exploiting the fact that customers only consider the lowest priced product available and the no purchase alternative when evaluating a product offering, the second approach computes the choice probabilities directly from the transaction data after accounting for the length of the booking horizon. The proposed approaches are detailed in Sub-sections 5.4.3 and 5.4.4.

5.4.3. Multinomial Logit Model

In various settings that relate to economics, psychology, marketing, transportation, etc., discrete choice models are used in an attempt to explore and explain how generic decision makers select a preferred alternative from a mutually exclusive, exhaustive and finite set of alternatives. Often, discrete choice models are derived based on the assumption that the decision maker is a rational individual who makes decisions in such a way that his overall utility is maximized. Since utilities are theoretically known to the decision maker only, they are typically modeled as linear functions of some observed attributes and, then, they are adjusted by random components to account for the influence of all other factors that affect the choice but are not present in the utility formulation. To this end, the utility U_{nj} the decision maker n obtains from alternative j can be expressed as: $U_{nj} = V_{nj} + \varepsilon_{nj} = \beta' \cdot x_{nj} + \varepsilon_{nj}$, $j \in S \cup \{0\}$, where V_{nj} and ε_{nj} are the observed and random components of the utility, S is the set of alternatives available to individual n , $\{0\}$ refers to the always present no purchase alternative, x_{nj} is a vector of observed attributes that identify alternative j , and β is a vector of parameters that need to be estimated. The vector of parameters β quantifies the relative contribution of each attribute to the observed utility component.

The utility decomposition together with the assumed behavioral decision rule facilitates the development of choice models that can be employed to compute the vector of parameters β . However, given that ε_{nj} is a random component that takes on unknown values, absolute statements relative to the decision maker's choice cannot be made. Rather, the choice can be explained only up to a probability that a current alternative is going to be selected. Thus, the probability the decision maker n chooses alternative j can be expressed as:

$$P_{nj} = P(U_{nj} > U_{nk}; \forall k, k \neq j) = P(\varepsilon_{nj} > V_{nk} - V_{nj} + \varepsilon_{nk}; \forall k, k \neq j). \quad (33)$$

In Equation (33) if the random components ε_{nk} are assumed to be independent and identically distributed and follow an Extreme Value Type 1 distribution then the MNL model is derived. Since the absolute magnitudes and the scales of the utilities are irrelevant, the MNL model requires ε_{nk} 's have the variances set to 1 to guarantee the identification of β parameters. With these assumptions in place, the probability expression shown in (33) simplifies and leads to $P_{nj} = e^{V_{nj}} / \sum_{k=1}^K e^{V_{nk}} = e^{\beta' \cdot x_{nj}} / \sum_{k=1}^K e^{\beta' \cdot x_{nk}}$, where V_{nk} (V_{nj}) is the observed component of the utility of alternative k (j), K refers to the cardinal of the decision maker's complete choice set $S \cup \{0\}$, x_{nk} (x_{nj}) is the vector of observed attributes that identify alternative k (j), β is a vector of unknown parameters, and e refers in this context to the exponential function. Since the observed utility of the no purchase alternative can be set to 0, the probability that decision maker n selects alternative j can be re-written as: $P_{nj} = e^{\beta' \cdot x_{nj}} / \left(\sum_{k \in S} e^{\beta' \cdot x_{nk}} + 1 \right)$.

The above probability expression establishes a relationship amongst the observed attributes x_n , the unknown parameters β and the probabilistic outcome P_{nj} . The intention here is to estimate parameters β in such a way that the probabilistic outcomes P_{nj} match as closely as possible the decision makers' observed choices. Stated otherwise, through several iterations that alter the values of parameters β , the aim is to maximize the likelihood that the observed choices are correctly identified by the choice model across all decision makers in the data sample. Thus, we want to determine the vector of parameters β that maximizes the likelihood function $L = \prod_{n=1}^N \left[\prod_{j \in S \cup \{0\}} \left(P_{nj}^{\delta_{nj}} \right) \right]$, where N provides the number of decision makers in the sample, P_{nj} is the probability of decision maker n selecting alternative j , and δ_{nj} is a dummy variable equal to 1 if alternative j is

the observed choice of decision maker n , 0 otherwise. However, to avoid running into numeric instabilities, in most applications that involve discrete choice models it is more common to maximize the logarithm of the likelihood function instead of the function itself. For the MNL model, the log likelihood function takes on the following expression:

$$LL = \log(L) = \log \left\{ \prod_{n=1}^N \left[\prod_{j \in C_n \cup \{0\}} (P_{nj}^{\delta_{nj}}) \right] \right\} = \sum_{n=1}^N \sum_{j \in C_n \cup \{0\}} \delta_{nj} \cdot \log(P_{nj}). \quad (34)$$

To be effective, all possible choice outcomes need to be represented in the log likelihood function depicted in Equation (34). However, the transaction level data that most service operators deal with does not fully support this requirement. In particular, the information about customers who request service but decide to not purchase is not retained since these customers do not directly contribute to the company's profit. Hence, the influence of these customers' decisions on the parameter estimates β is completely overlooked. Since the booking horizon can be discretized so finely that a decision period could be associated with at most one arrival, the inability to discriminate between time periods with no arrivals at all and those with arrivals but not observed purchases makes the classic MNL estimation techniques inapplicable to this situation. Instead, an alternative approach pioneered by Talluri and van Ryzin ([14], [35]) that builds on the expectation-maximization (EM) technique of Dempster, et al. [66] is employed.

In this framework, D , P , and $\bar{P} = D - P$ are used to denote the entire set of decision periods, the set of periods in which a purchase is observed, and the set of periods for which no-purchase transactions are reported, respectively. The MNL model is assumed to appropriately describe how customers make purchasing decisions. Furthermore, $j(t)$ is considered to represent the product the customer arriving in time period $t \in P$ purchases from the choice set $S(t)$ he is offered for consideration. With these notations, the complete likelihood function for the entire booking horizon becomes:

$$L = \prod_{t \in P} [\lambda \cdot P_{j(t)}(x, \beta, S(t))] \cdot \prod_{t \in \bar{P}} \left\{ [\lambda \cdot P_0(x, \beta, S(t))]^{a(t)} \cdot (1 - \lambda)^{(1-a(t))} \right\}, \quad (35)$$

where λ is the unknown arrival rate of customers assumed constant over the booking horizon, $P_{j(t)}(x, \beta, S(t))$ is the purchasing probability associated with the purchase period t and choice made $j(t)$, $P_0(x, \beta, S(t))$ is the no purchase probability associated with the no purchase period t , $a(t)$ is an indicator variable equal to 1 if a customer's request for a product arrives in period t , 0 otherwise. In the expression depicted in Equation (35), the product over all periods $t \in P$ computes the likelihood that the observed outcomes are correctly replicated by the behavioral model employed. This likelihood is then adjusted through the use of the second product to account for the influence of the arrival - no purchase mechanism that characterizes the time periods in \bar{P} . In particular, in the second product, it is acknowledged that a no purchase can be caused by a customer requesting a product and deciding not to buy (i.e., $\lambda \cdot P_0(x, \beta, S(t))$), or, by a customer not requesting service at all (i.e., $(1 - \lambda)$). Since these instances exclude each other, the presence of the exponents $a(t)$ and $(1-a(t))$, respectively, guarantees the independence of the two events.

In Equation (35), besides λ and β , the $a(t)$'s for time periods $t \in \bar{P}$ are also unknown. However, since their expected values can be computed, one can apply the EM method to recommend improved estimates for all unknowns. Thus, during the Expectation step, the expected values for all $a(t)$'s, $t \in \bar{P}$, are computed using a current vector of parameter estimates λ and β . Then, during the Maximization step, these values help recommend refined λ and β values by maximizing the logarithm of the likelihood function shown in (35). The adjustment process repeats until convergence, or, equivalently, until the difference in two consecutive sets of parameter estimates λ and β does not exceed a predetermined threshold value. Lastly, once an optimal set of parameter estimates λ and β is computed, the choice probabilities $P_j(S)$ for all alternatives

$j \in S$ and $S \subseteq N$ can be accordingly estimated using the MNL alternative substitution patterns.

5.4.4. Alternative Approach to Multinomial Logit Model

The complexity of the EM - MNL approach to computing the choice probabilities $P_j(S)$ may deter service providers from widely accepting it. Indeed, to date, there are not commercially available software packages that incorporate the customer purchasing behavior into the framework required by the EM methodology. Thus, if such an approach is to be considered, then all the algorithms that support this initiative need to be developed in house or outsourced to consultant companies. The lack of in house qualified personnel and/or the forbidden prices quoted by consultants may lead, however, to this approach being considered prohibited. In this section, an alternative approach to computing choice probabilities $P_j(S)$ in a lowest fare available environment is devised. The new approach is simpler, yet intuitive, and successfully exploits the purchasing patterns observed in a price sensitive environment.

Consistent with all previous assumptions, the service provider is considered to sell the same perishable product to multiple customer segments which accept a predetermined differential pricing schema. For a judiciously designed such schema, all customers requesting service should purchase the lowest price point if it is made available. Thus, it can be reasonably asserted that when the complete set of price points is offered, the probability to purchase is equal to one. In this case, the probabilities to purchase and purchase the lowest price point, respectively, coincide. When the lowest price point is removed from the offered set, some customers will purchase the new lowest priced product available while others will divert and look for alternative options. The increase in the lowest price point available leads to lost demand, or, equivalently, to a nonzero probability of no purchase. If sales figures associated with offering the complete and incomplete product sets during the entire booking horizon were known, then the

probability to not purchase associated with the latter could be computed as

$P_{NoP}^{S(i)} = \left(N_{min(S(c))} - N_{min(S(i))} \right) / N_{min(S(c))}$, where $S(c)$ and $S(i)$ are the complete and incomplete product sets, respectively, and $N_{min(S(c))}$ and $N_{min(S(i))}$ refer to the sales observed at the lowest price products available in sets $S(c)$ and $S(i)$. Obviously, the probability to purchase which would be identical to the probability of purchasing the lowest priced product available would equate to $(1 - P_{NoP}^{S(i)})$.

Unfortunately, the required sales figures are typically not available since service providers tend to control the availability of their price points during the booking horizon. However, estimates for these numbers can be obtained by projecting the sales at different price points observed during time periods $(t \leq T)$ over the entire booking horizon T . Since time intrinsically impacts the observed sales, time series forecasting techniques are employed to project the sales over the entire length of the booking horizon. Hence, the probability of no purchase associated with set $S(i)$ can be re-written as:

$$P_{NoP}^{S(i)} = \left[E\left(N_{min(S(c))}\right) - E\left(N_{min(S(i))}\right) \right] / E\left(N_{min(S(c))}\right),$$

where $E(\dots)$ refers to the expected sales over the entire booking horizon at the lowest priced products available in sets $S(c)$ and $S(i)$, and all other entities keep their previous meanings. In this framework, the methodology applies to all incomplete sets as long as product classes are cut off in ascending order of their prices. For example, when the RM system recommends closing the lowest priced product available in set $S(i)$, the probability of no purchase associated with the resulting set $S(j)$ can be easily estimated by replacing the entities that refer to $S(i)$ with those corresponding to $S(j)$.

The computation of probabilities $P_{NoP}^{S(i)}$ for all incomplete product offerings $S(i)$ allows one to specify probabilities $P_j(S(i))$ required by the choice-based capacity control

techniques. To this end, replacing $S(i)$ with S for conciseness, probabilities $P_j(S)$ equal $(1 - P_{NoP}^S)$ if j is the lowest priced product in set S , and 0 otherwise.

5.5. Experimental Setup

In this study, the case of a service provider that sells three products at \$100, \$200, and \$300, respectively, is examined. The products are considered identical in all regards but the price. The booking horizon at this service provider is assumed to consist of 300 time units. Customer arrivals during the booking horizon are considered to follow a homogeneous Poisson process with a rate parameter λ of 0.5, or, equivalently, a mean of 150. Customers are assumed to arrive with a willingness to pay that takes on one of three possible values: \$125, \$225, or \$325, respectively. For simplicity, each of these values is considered to be equally likely to be assigned to an arriving customer. In such an environment, a customer is assumed to always make a purchase against the lowest priced product available if and only if the customer's willingness to pay exceeds this product's price.

To closely replicate the actions taken by a real RM system, the availability of offered products is restricted during the booking horizon. In this context, an initial set of nested booking limits, set arbitrarily, is dynamically updated as more information about product demand becomes available. For example, without any prior knowledge about the demand, the revenue manager at the service provider may opt to control his capacity by using 50 and 100, respectively, as his initial booking limits. This means that the \$100 product becomes unavailable when the fiftieth unit of capacity is sold. Similarly, when the one hundredth unit of capacity is purchased, the \$200 product becomes unavailable. The \$300 product stays available until the entire capacity sells out or the end of the booking horizon is reached. A current set of booking limits stays in place until a batch of 10 consecutive arrival streams is processed. Then, the unconstrained demand associated with these arrivals is appended to that that has previously been considered, and new

booking limits are computed for the next arrival batch based on the aggregated figures. The process repeats several times - ten times in this study - until the booking limits do not significantly change from iteration to iteration. To mitigate the spiral down effect (Boyd, et al. [81], Cooper, et al. [82]) that accompanies this iterative process only the nonzero unconstrained demand levels are used to recommend refined booking limits. The zero demand levels reported when a product class is permanently closed and/or when it is open but all requests materialize in sales in lower classes are ignored as using them may unjustifiably underestimate the true demand for that product class.

In the example above, dynamically updating the booking limits may result in a recommended pair of 59 and 132, respectively. The evolution of booking limits across all arrival batches when the process repeats multiple times is shown in Figure 19.

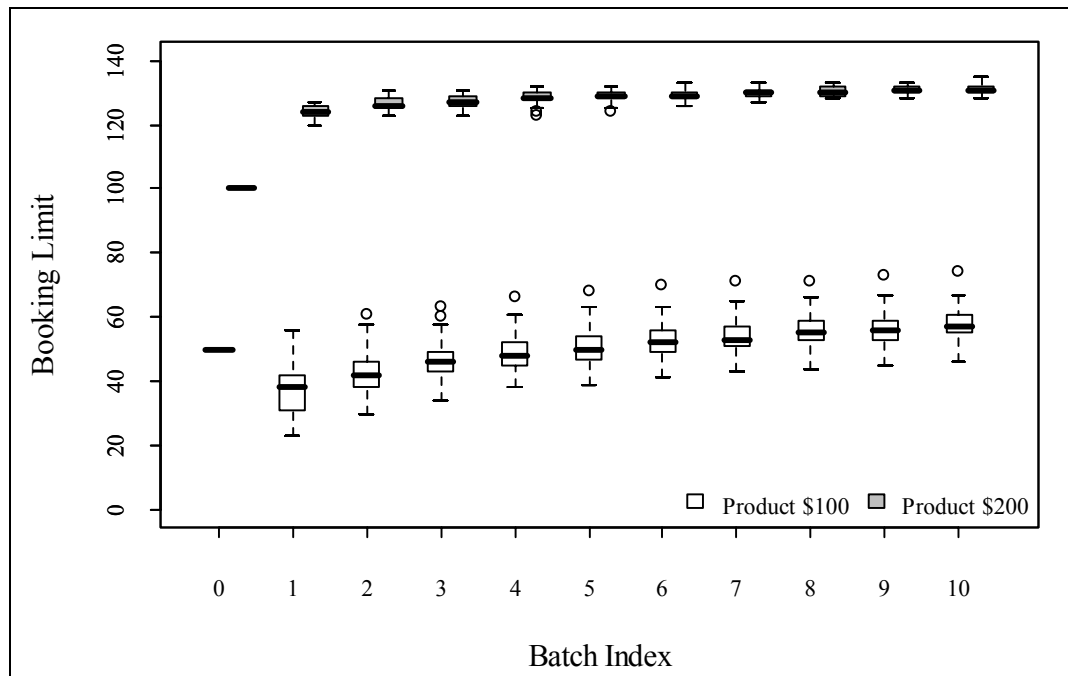


Figure 19 Box-Plots of the Booking Limits by Arrival Batch (Capacity 140; 50 Repetitions)

5.6. Capacity Control Policies

The booking limit refinement procedure described in Section 5.5 mimics the behavior of a newly initialized RM system that regularly adjusts to better fit changing market

conditions. As more information becomes available, the system matures and tends to respond more uniformly to requests for service. In this study, product availability during the transition period is controlled using the EMSR-b heuristic. The transaction level data used to compute the refined protection levels, enriched with the time of booking availability information, is utilized to determine the competing choice-based control policies. The procedures that allow the computation of these policies are detailed next.

5.6.1. EMSR-b Heuristic

Finding the EMSR-b control policy that best describes the studied lowest price available environment intrinsically relates to the process of unconstraining product demand. The theoretical concepts that lead to the recommended EMSR-b protection levels are extensively discussed in Sub-section 5.4.1. As a result, we focus on how unconstraining the demand should be done to support an improved system performance and, subsequently, on how the two processes interconnect.

Consistent with the RM literature, the unconstrained product demand distributions are assumed to be normally distributed. Their parameters are successively refined based on the demand that is processed once a complete batch of customer arrival streams materializes. Within an arrival batch, the times products become unavailable as well as the time sequences associated with the incremental product demands are recorded for each individual arrival stream. To illustrate, during the initial iteration, the fiftieth unit of capacity, or, equivalently, the fiftieth \$100 product, could be sold during the 88th time unit of the booking horizon. Similarly, the one hundredth unit of capacity, or, equivalently, the fiftieth \$200 product, could be sold during the 233rd time unit of the booking horizon. For this hypothetical example, a possible plot of the incremental product demands is shown in Figure 20.

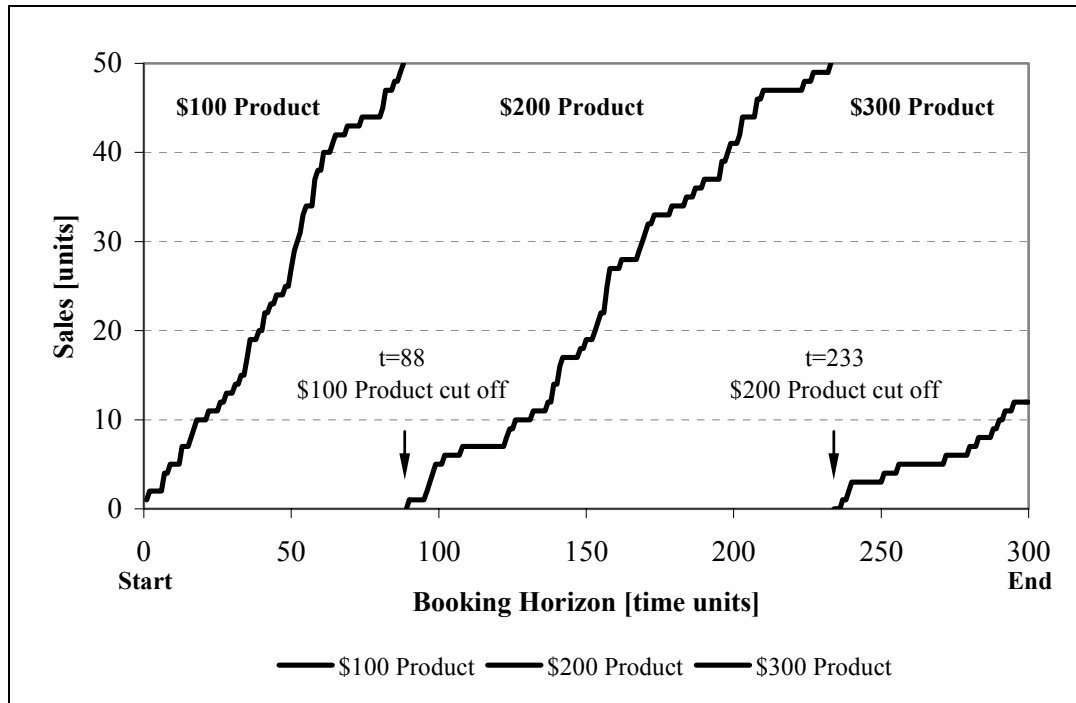


Figure 20 Incremental Product Demand Plots for a Hypothetical Arrival Stream

The time-sales information at the arrival stream level facilitates the computation of the unconstrained product demand, that is, the demand for individual products that would have been observed in the absence of any capacity control restrictions. While several methods can be used to determine the unconstrained demand levels (i.e., averaging methods (Weatherford and Pölt [85], Zeni [86]), booking profile (Pölt [87], Zeni [88]), expectation-maximization (Zeni [86], Zeni [88]), projection detruncation (Weatherford and Pölt [85], Zeni [86], Zeni [88])), the double exponential smoothing (DES) (Crystal, et al. [89]) is preferred because of its reported accuracy and implementation simplicity. The use of the DES method is also supported by the shapes of the incremental product demand plots which display significant trends but no seasonality. Comprehensive reviews of the DES method are provided in forecasting textbooks such as Armstrong [90], Brockwell and Davis [91] and Chatfield [92].

Continuing the example illustrated in Figure 20, the unconstrained demand levels as forecasted by the DES method are shown in Figure 21. As depicted on the right side of the figure, the DES method suggests that 164 units of capacity would have been sold at

\$100 if there had not been any capacity controls in place. Similarly, the method indicates that 66 and 12 units of capacity would have been probably sold at \$200 and \$300, respectively, if the service provider had not restricted the availability of its capacity.

The DES technique, repeatedly applied to all simulated arrival streams associated with an arrival batch, helps compute the parameter estimates of the updated unconstrained demand distributions to be used to set the values of the next booking limits. To illustrate, for a hypothetical case of two consecutive arrival batches (i.e., the first two of the 10 possible batches), each consisting of 10 arrival streams, the successive application of the DES method leads to the unconstrained parameter estimates depicted in Table 16. In this example, the computation of the second pair of booking limits follows the steps described in Sub-section 5.4.1 where, for convenience, the protection levels are converted to the corresponding booking limits. Finally, iterating over all arrival batches, the recommended EMSR-b control policy is determined. As shown in Table 17, this policy suggests that 9 units of capacity should be reserved for customers purchasing the \$300 products, 90 units should be reserved for customers requesting the \$200 and \$300 products, and the remaining capacity, or, 50 units can be sold at \$100.

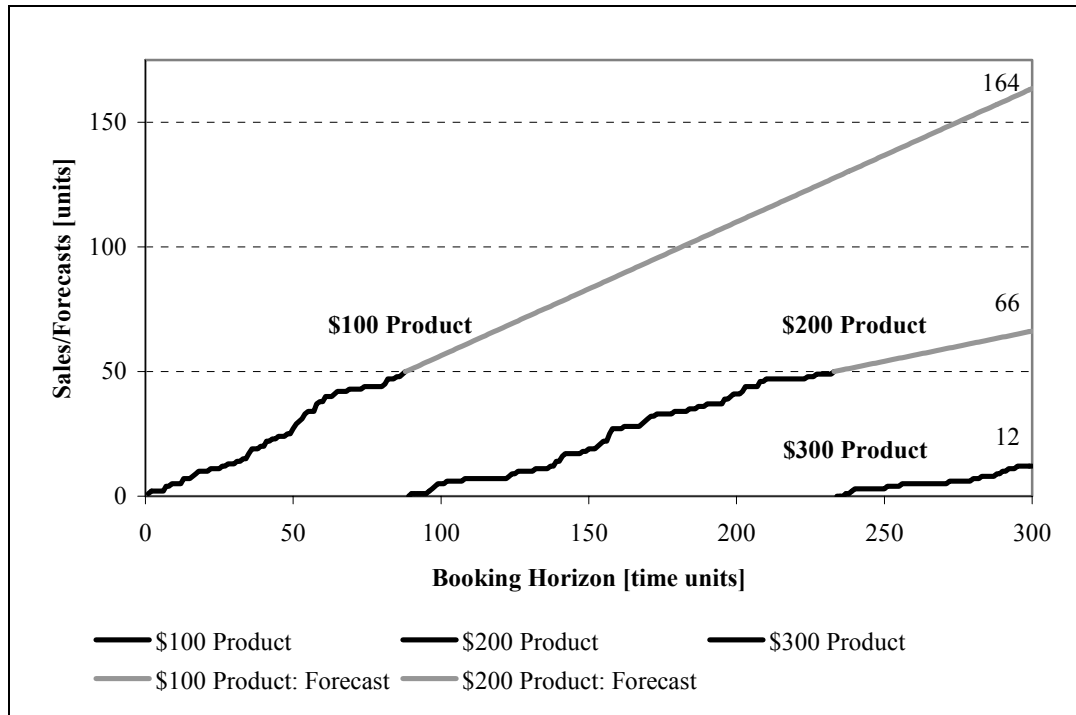


Figure 21 Unconstrained Product Demand Levels for a Hypothetical Arrival Stream

Table 16 Parameter Estimates for the Unconstrained Product Demand Distributions

Product	Price	Available Products	Unconstrained Product Demand Distributions ^{a)}			
			Iteration 1 ^{b)}		Iteration 2 ^{c)}	
			Mean [units]	Std. Dev. [units]	Mean [units]	Std. Dev. [units]
\$100 Product	\$100	All	168.8	57.6	170.2	48.4
\$200 Product	\$200	\$200, \$300	78.3	17.5	73.7	14.8
\$300 Product	\$300	\$300	18.5	6.1	17.3	7.1

^{a)} Capacity: 140 units; Number of arrival streams per batch: 10. ^{b)} Requests are accepted / denied based on the initial booking limits of 50 and 100. ^{c)} Requests are accepted / denied based on the updated booking limits of 41 and 121.

Table 17 EMSR-b Heuristic - Nested Protection Levels

Product	Price	Unconstrained Product Demand Distributions		Protection Levels	Booking Limits
		Mean [units]	Std. Dev. [units]		
\$300 Product	\$300	11.9	8.9	9	cap = 140
\$200 Product	\$200	76.7	11.7	90	131
\$100 Product	\$100	174.1	60.3	cap = 140	50

5.6.2. Choice-Based Capacity Control Technique

As discussed in Section 5.4, the choice-based capacity control techniques employ customer purchasing behavior models to recommend the optimal control policies. The two approaches used to provide behavioral inputs to the optimization routines are detailed next. In both cases, a brief description of how the transaction level data is manipulated to support the estimation of these inputs accompanies the exposition.

The sequence of arrival batches used to compute the EMSR-b protection levels assists in creating the input data sets for the EM method. In this context, the products that are available during each time period of the booking horizon are recorded independent of whether or not arrivals are being observed. Thus, the data requirements needed to estimate an MNL model of customer purchasing behavior using transaction level data are all met. In an environment in which the price is the only driver of the purchasing decision, the EM method leads to a possible set of parameter estimates that are depicted in Table 18. In this case, β_{price} quantifies the average impact that a one dollar increase in price has on the observed utilities, β_0 , associated with the nonzero price alternatives only, is an additive utility constant that guarantees that the no purchase option stays the reference alternative (i.e., $U_{No\ Purchase} = 0.0$), and λ is the recovered customer arrival rate. The vector of parameters $\beta = (\beta_0, \beta_{price})$ together with the MNL alternative substitution patterns allows one to compute the choice probabilities $P_j(S)$ associated with any offered set S . For all product sets that a customer may face in our experiment, these probabilities are shown in Table 19. Consistent with the experimental assumptions, when a complete product set is offered, an arriving customer considers the purchase of the minimum priced product available with a probability close to 1.0. Similarly, when the \$100 product is removed from the offered set, an arriving customer is considered to purchase the newly lowest priced product available with a probability that also approaches 1.0. While a high probability to purchase the \$200 product is to be

expected, its extreme value may be attributed in part to how the MNL model redistributes the choice probabilities when some alternatives are removed from the choice sets. However, unless other more sophisticated discrete choice models are employed, to date there are no techniques that account for the limited MNL substitution patterns when recommending the required choice probabilities $P_j(S)$. Lastly, when the \$300 product is offered alone, customers appear to be more likely to not purchase rather than purchase at the proposed price point.

Table 18 MNL Parameter Estimates Computed using the EM Method

Parameters ^{a)}	Estimates	Std. Err.	Est./S.E.	Sig.
β_{Price}	-0.112	0.017	-6.68	0.00
β_0	33.189	5.013	6.62	0.00
λ	0.330	-	-	-

^{a)} Capacity: 140 units. The estimation data set consists of purchase/no purchase observations from all arrival streams and all arrival batches processed.

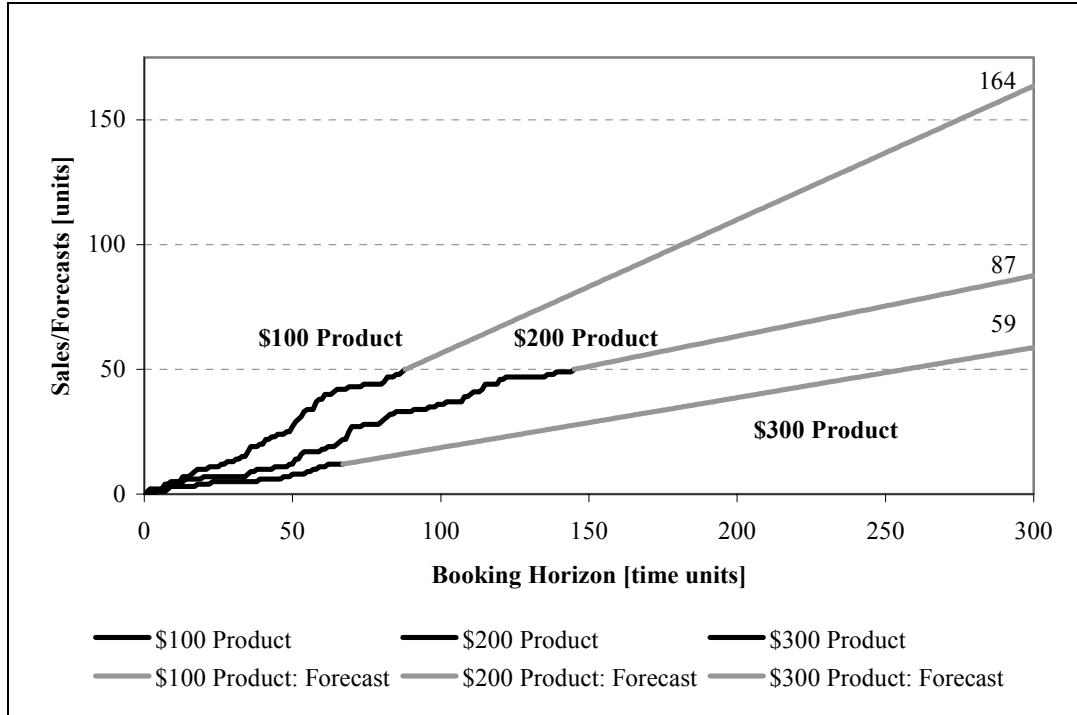
Table 19 MNL Choice Probabilities

Choice Probabilities	Offered Product Set ^{a)}		
	S_1	S_2	S_3
$P_{No\ Purchase}$	0.00000	0.00002	0.60133
$P_{\$100}$	0.99999	-	-
$P_{\$200}$	0.00001	0.99997	-
$P_{\$300}$	0.00000	0.00001	0.39867

^{a)} Capacity: 140 units. Product sets S_k , $k=1, 2, 3$, consist of products $\{\$100, \$200, \$300\}$, $\{\$200, \$300\}$ and $\{\$300\}$, respectively.

As opposed to how the EM behavioral outcomes are computed, the alternative choice-based approach requires the estimation of the demand that the service provider would have experienced had it offered exclusively certain product sets over the entire booking horizon. To accomplish this, the DES method consistently projects now the observed product sales over a forecast horizon equal to the booking horizon. In this case, the unconstrained demand levels dependent upon the offered product sets are shown in Figure 22 where, for convenience, the incremental demand plots share the same origin. As depicted on the right side of the figure, the DES method suggests that 164 units of capacity would have been sold at \$100 if the product set consisting of all products had

been offered during the entire booking horizon. Similarly, the method indicates that 87 and 59 units of capacity would have been purchased at \$200 and \$300, respectively, if the product sets consisting of the \$200 and \$300 products and the \$300 product, respectively, had been exclusively offered during the booking horizon.



**Figure 22 Unconstrained Product Demand Levels for a Hypothetical Arrival Stream:
Demand Dependent upon the Offered Product Sets**

The demand estimates presented in Figure 22 measure customers' willingness to pay for service and should not be sensitive to the capacity controls in place at the time of their estimation. As a result, requests arriving during the booking horizon are accepted or rejected based on booking limits that allow sales to be made in each of the offered product classes. In all situations investigated in this study, the employed booking limits replicate those used to initialize the iterative process described in Section 5.6.1. In this case, given that a repetitive approach is not justified, the final parameter estimates for the required unconstrained demand distributions are computed based on a unique batch of 100 independent arrival streams. For a hypothetical such arrival batch, the unconstrained

parameter estimates of the demand distributions are shown in Table 20. Of all estimates, those associated with the \$300 product may, at times, be subject to overestimation bias due to the unfortunate situation that combines a short observed booking period with a long forecast horizon. Wherever apparent, these instances should be removed to maintain the parameter estimates unbiased.

Table 20 Parameter Estimates for the Unconstrained Product Demand Distributions:

Demand Dependent upon the Offered Product Sets

Product	Price	Available Products	Unconstrained Product Demand Distributions ^{a)}	
			Mean [units]	Std. Dev. [units]
\$100 Product	\$100	All	173.7	38.7
\$200 Product	\$200	\$200, \$300	107.2	24.2
\$300 Product	\$300	\$300	52.3	26.8

^{a)} Capacity: 140 units; Number of arrival streams: 100; Requests are accepted / denied based on the initial booking limits of 50 and 100.

The assumption that governs the assignment of customer's willingness to pay together with the suggested price structure leads to an arriving customer always purchasing the \$100 product if a complete product set is made available to him. Thus, a probability of purchase of 1 can reasonably be associated with a request that arrives when a complete product set is offered. In such a situation, the hypothetical number of \$100 products that would have been sold if the complete product set had been offered during the entire booking horizon is provided by the DES technique (e.g., 173.7 in Table 20). When the \$100 product is removed from the offered sets, customers perceive this action as a threat and respond by purchasing less of the available capacity. The amount customers would have purchased had they been offered the set consisting of the \$200 and \$300 products is also given by the unconstraining method (e.g., 107.2 in Table 20). The difference in the two forecasts relative to the first of them can be viewed as the resistance of customers to price increases and, subsequently, interpreted as the probability of no purchase associated with offering product set (\$200, \$300). Similar reasoning applies to computing the probability of no purchase corresponding to offering product set (\$300).

Following the methodology described in Section 5.4.4, the choice probabilities $P_j(S)$ associated with any offered product set S can take on the values depicted in Table 21.

Table 21 Choice Probabilities - Alternative Approach

Choice Probabilities	Offered Product Set ^{a)}		
	S_1	S_2	S_3
$P_{No\ Purchase}$	0.000	0.380	0.700
$P_{\$100}$	1.000	-	-
$P_{\$200}$	0.000	0.620	-
$P_{\$300}$	0.000	0.000	0.300

^{a)} Capacity: 140 units. Product sets S_k , $k=1, 2, 3$, consist of products $\{\$100, \$200, \$300\}$, $\{\$200, \$300\}$ and $\{\$300\}$, respectively.

The two sets of choice probabilities $P_j(S)$ enter the objective function of the dynamic program formalized in Equation (32) and assist in computing the optimal choice-based capacity control policies. To illustrate the latter process, let's consider the case of the choice probabilities $P_j(S)$ depicted in Table 21. The total probability of purchase associated with offered sets S_1 , S_2 , and S_3 equates to 1.00, 0.62, and 0.30, respectively. Similarly, the total expected revenues that can be obtained from offering sets S_i , $i \in \{1, 2, 3\}$, equal \$100, \$124, and \$90, respectively. Substituting these values in Equation (32) and maximizing over all offered sets S_i lead to an optimal total expected revenue $V_1(x)$ of \$62. In this case, given that products completely lose their values after they expire and the displacement cost of a unit of capacity is zero at the time of spoilage, the maximization problem reduces to $V_1(x) = \max_{S_i} \{\lambda \cdot R(S_i)\}$, or, $V_1(x) = \$62$. Similar reasoning allows to recursively compute the optimal total expected revenue $V_t(x)$ for all time units t of the booking horizon and all remaining capacity levels x . Table 22 provides an illustration of how the $V_t(x)$ values may evolve on both directions considered.

Table 22 Optimal Total Expected Revenue $V_t(x)$

Time to Spoilage [Time Units]								
...	6	5	4	3	2	1	0	
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0
198.24	180.28	159.15	134.30	104.78	62.00	62.00	0.00	1
303.19	268.56	227.86	180.04	124.00	62.00	62.00	0.00	2
353.97	302.48	246.15	186.00	124.00	62.00	62.00	0.00	3
369.27	309.43	248.00	186.00	124.00	62.00	62.00	0.00	4
371.82	310.00	248.00	186.00	124.00	62.00	62.00	0.00	5
372.00	310.00	248.00	186.00	124.00	62.00	62.00	0.00	6
								...

Remaining Capacity

To illustrate the potential benefits associated with employing choice-based controls, let's assume that the choice-based control policy has been computed and 11 time units prior to product spoilage the service provider has 5 more units of capacity in inventory. In addition, during the remaining of the booking horizon, requests for products are assumed to arrive as shown in the upper part of Figure 23. When the first of them materializes, the optimal choice-based policy recommends the display of set S_1 which consists of all three products. Since the willingness to pay corresponding to this request is above the minimum price in the display set, a sale is made against the lowest priced product available (i.e., \$100). Given that the remaining capacity is updated once a sale is made, the second customer is shown set S_2 which consists of the two most expensive products. Since the customer's willingness to pay does not exceed the minimum price in the display set, the second product request is in this case denied. Similar reasoning leads to the service provider accepting the third and the fourth requests with sales being reported against the \$200 product and denying the last of its customers. During the time period considered, the service provider earns \$500 in revenues which translates into an incremental revenue gain of \$200 relative to the amount it would have reported had it followed the recommended EMSR-b policy (see Table 17). While the magnitude of the results is specific to this particular example, the findings support the hypothesis that exploiting choice behavior may lead to improved revenue performance.

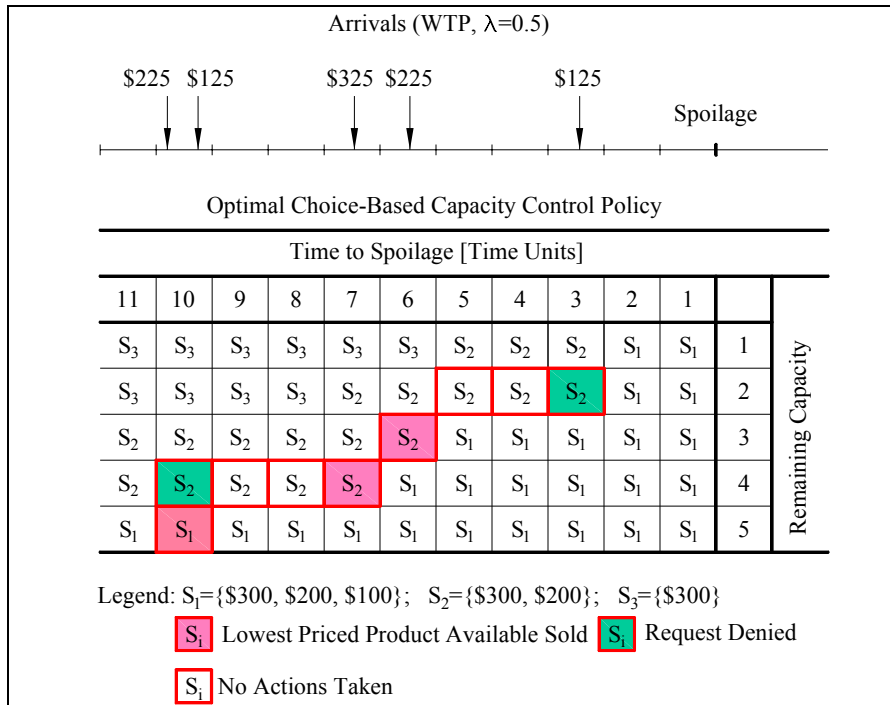


Figure 23 Optimal Choice-Based Capacity Control Policy - Example

5.7. Revenue Performance Evaluation

To compare the performance of the competing policies, multiple customer arrival streams are generated and the controls of each policy, computed at capacity levels ranging from 90 to 150 in steps of size 10, are applied. The rules for accepting/denying an arriving request are detailed in the introductory statements of Section 5.5. The revenues associated with each of the methods considered for a capacity level of 100 units and 50 revenue evaluation arrival streams are shown in Table 23. The revenue statistics are accompanied by other non-monetary, system performance measures that provide an overall image of the system's utilization.

Table 23 Revenue Performance and other Performance Measures

Performance Measure ^{a)}	EMSR-b	Ch-B ^{b)}	Ch-B EM ^{c)}
1. Revenues			
Average	\$19,578	\$19,772	\$19,802
Std. Dev.	\$1,208	\$1,369	\$1,400

Table 23 Continued

Performance Measure ^{a)}	EMSR-b	Ch-B ^{b)}	Ch-B EM ^{c)}
2. Sold Capacity			
Average	95.7	96.6	96.5
Std. Dev.	4.6	4.7	4.6
3. Unsold Capacity			
Average	4.3	3.4	3.5
Std. Dev.	4.6	4.7	4.6
4. Number of Requests			
Average	150.2	150.2	150.2
Std. Dev.	12.9	12.9	12.9
5. Denied Requests			
Average	51.3	51.4	52.3
Std. Dev.	7.9	8.5	9.0
6. Unprocessed Requests			
Average	3.2	2.2	1.4
Std. Dev.	7.2	3.5	2.5

^{a)} Capacity: 100 units; Revenue Evaluation Arrival Streams: 50.

^{b)} Ch-B: Alternative Choice-Based Control Policy.

^{c)} Ch-B EM: EM Choice-Based Control Policy.

In an attempt to ensure that the relative revenue performance of the methods investigated is consistent across independent trials, the processes detailed in Sections 5.5 - 5.6 are repeated multiple times at each level of capacity considered. The spread of the revenues reported for 50 such repetitions is shown in the box-plots depicted in Figure 24. At low capacity levels, irrespective of the control methods employed, 50% of the reported revenues are within bands of reduced width. In contrast, as the available capacity increases, so does the variability in revenues, the 50% revenue bands more than doubling their sizes at the highest levels of capacity.

The information presented in Figure 24 , summarized over all capacity levels investigated, is shown in Figure 25. In its upper panel, at each capacity level, the average revenues over all considered repetitions are displayed. This information is supplemented in the lower panel with the computed revenue differences expressed as percentages of the EMSR-b revenues. Consistent across all capacity levels, choice-based control policies show superior revenue performance when the lowest open fare model is employed. Moreover, while somewhat different at low levels of capacity, these policies report

similar earnings at medium and high capacity levels. Statistically, these findings are shown in Table 24 which compiles the results of all relevant two-sample paired t-tests.

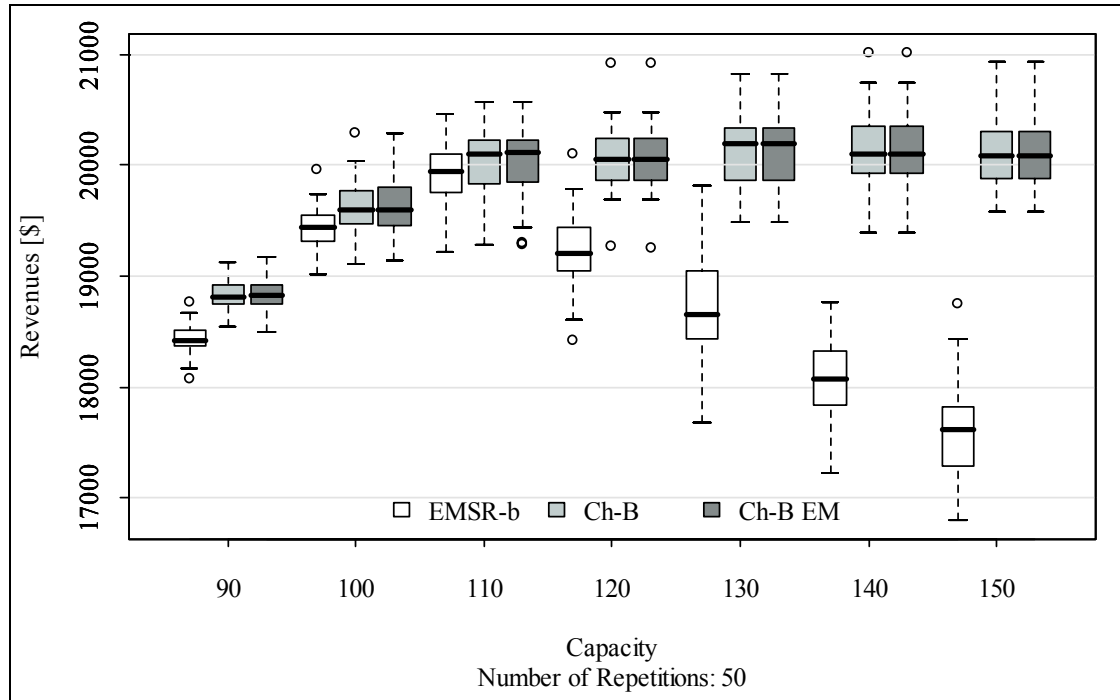


Figure 24 Box-Plots of Revenues Reported by Competing RM Methods

Table 24 Revenue Performance Comparison - Two-Sample Paired t-test Results

	Capacity [units]						
	90	100	110	120	130	140	150
EMSR-b vs. Ch-B (Alternative Hypothesis: Ch-B performs better than EMSR-b)							
Revenue Difference	-400.2	-185.0	-109.8	-833.6	-1372.7	-2034.4	-2521.5
t statistic	-28.2	-19.4	-7.5	-29.6	-35.2	-61.4	-57.0
df	49	49	49	49	49	49	49
P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Ch-B vs. Ch-B EM (Alternative Hypothesis: Ch-B EM performs better than Ch-B)							
Revenue Difference	-14.1	-8.4	-5.8	0.3	-0.2	0.0	0.0
t statistic	-2.6	-2.9	-4.0	0.4	-0.8	-	-
df	49	49	49	49	49	49	49
P-value	0.01	0.00	0.00	0.66	0.21	-	-

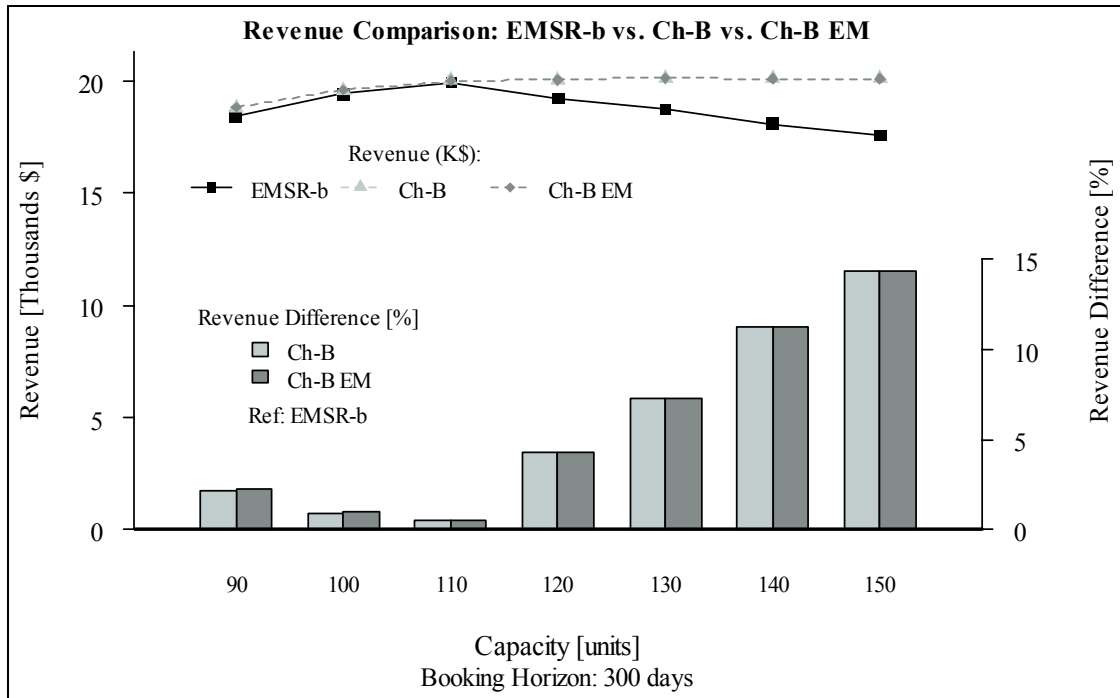


Figure 25 Revenue Comparison: EMSR-b vs. Ch-B vs. Ch-B EM

The revenue differences depicted in Figure 25 can be explained by how the competing control policies manage the capacity levels considered. At low levels, EMSR-b closes the \$100 product class and accepts requests for the higher priced classes only. As the capacity increases, EMSR-b opens the \$100 product class and proportionally accepts more requests for this product. In contrast, the choice-based policies keep the \$100 product class closed at all levels of capacity and consistently offer customers product sets consisting of the \$200 and \$300 products, and the \$300 product, respectively. In this context, as the capacity increases, so does the likelihood that an arriving customer is shown a product set consisting of both the \$200 and \$300 products.

In Figure 25, for capacity constrained regimes, the choice-based policies close the lower price product classes more aggressively than the EMSR-b to better exploit customers' willingness to pay. At corresponding load factors/occupancies similar in magnitude, the revenue increases from the higher price product classes decisively contribute to the slight revenue superiority of the choice-based capacity controls. An illustrative example of such a situation is provided in Figure 26 which shows the sales of

the \$200 and \$300 products at a capacity of 90 units and across 50 distinct arrival streams. With reported average load factors/occupancies around 98%, the choice-based policies sell, on average, 2.0 more units at \$300 and, thus, they tend to outperform the competing EMSR-b technique.

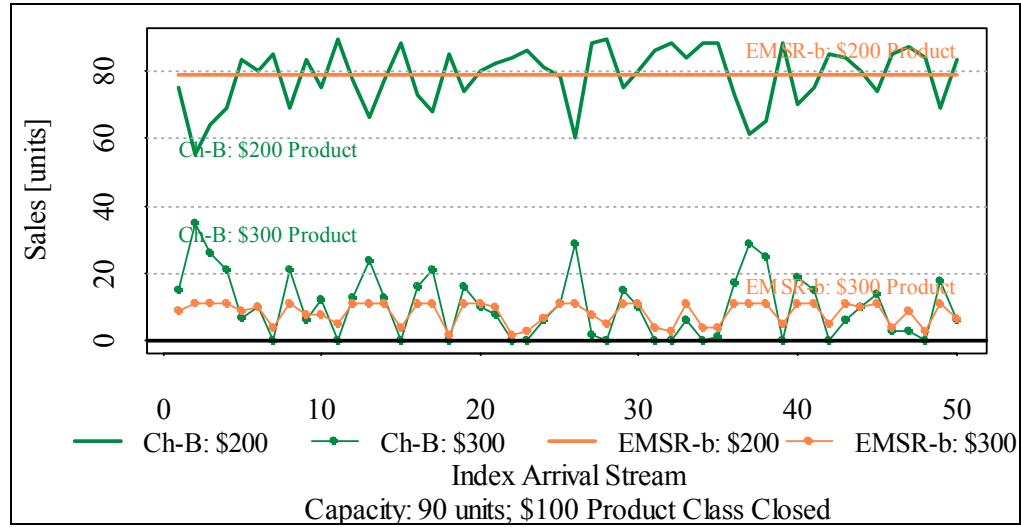


Figure 26 Sales by RM method and Product Price - Capacity 90 units

As the available capacity increases, the sharp difference between the sales of the \$300 product as reported by EMSR-b and the choice-based policies diminishes with both methods reporting insignificant sales for this product class. In this context, since the booking limit for the \$200 product class is rarely reached and the \$100 product class is always closed, the two competing RM methods yield similar revenues. This is best observed at a capacity level of 110 units, where the sales profiles for the \$200 product are similar. In Figure 27, the upper profiles do not perfectly match due to the fact that EMSR-b, at times, reaches its booking limit for the \$200 product class. In this case, the decision to keep selling at \$200 when there is indeed scarce demand for higher priced products leads to the negligible revenue superiority shown by the choice-based control policies (see Figure 25 and Table 24).

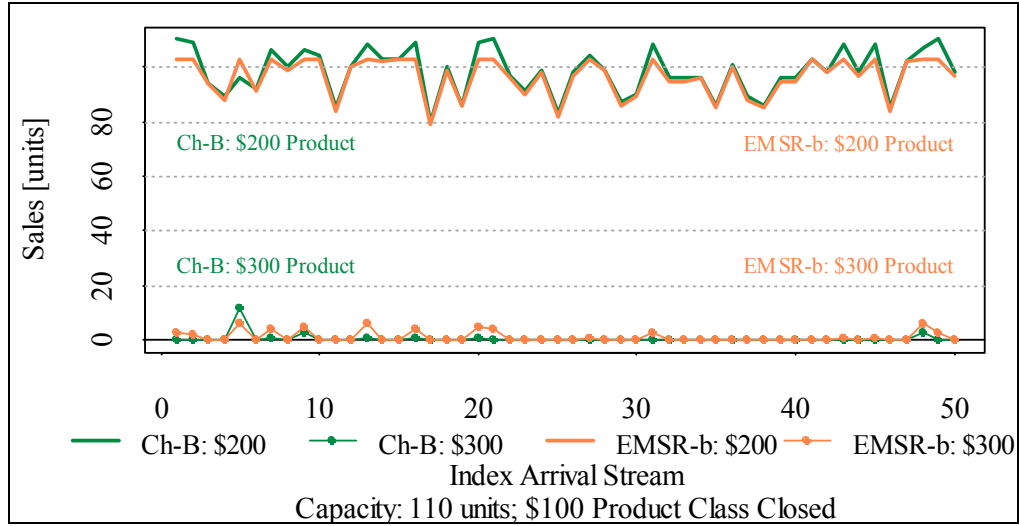


Figure 27 Sales by RM method and Product Price - Capacity 110 units

Moving to the unconstrained capacity regimes, the issues identified above become even more severe. In particular, while both methods rarely sell the \$300 product class, EMSR-b accepts an increasing number of \$100 requests even if the corresponding products could be easily sold at \$200. The choice-based control policies, on the other hand, recognize the potential for higher revenues and keep the \$100 product class closed during the entire booking horizon. An example of such a situation is presented in Figure 28 which shows the number of products sold at \$100 and \$200, respectively, for a capacity level of 120 units. In this case, the average offset of 14.0 units observed in between the \$200 product sales profiles compensate for the average loss of sales of 7.0 units associated with using the optimal choice-based policy. As the available capacity increases, the reported revenue superiority of the choice-based controls increases as well (see Figure 25).

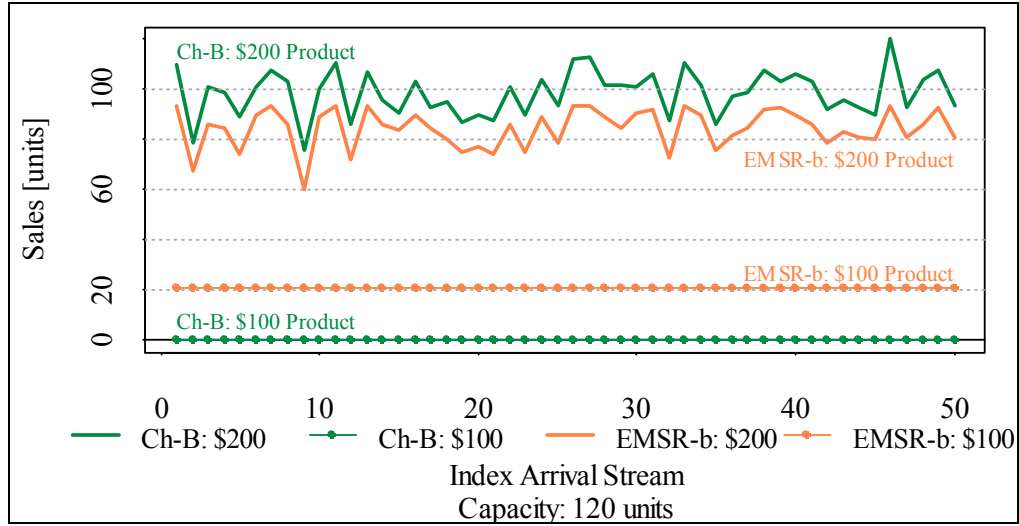


Figure 28 Sales by RM method and Product Price - Capacity 120 units

5.8. Summary

The current simulation study investigates the impact that the type of the employed RM control has on the performance of service providers that operate in price sensitive business environments. In this context, the revenue performance of traditional and choice-based capacity control policies is compared in an attempt to determine whether or not one of these techniques consistently outperforms the others. The simulation results suggest that the benefits service providers could experience from computing the RM controls to account for customer purchasing behavior are significant. Across all capacity levels investigated, the incremental revenue gains associated with the choice-based methodologies range between 1.0% and 14.0%.

Throughout this study, the reported incremental revenue gains are shown to be stable with respect to how the behavioral inputs are computed. In spite of the apparent mismatch between the choice probabilities recommended by the competing choice-based approaches, the corresponding control policies show similar yield. This recommends the approach we propose as a feasible alternative to the more computationally involved Expectation-Maximization technique. In addition, the choice-based control policies demonstrate consistent performance even in unconstrained capacity regimes where the

traditional EMSR-b technique reports weak earnings. While these results are encouraging, they are nevertheless the outcome of processes that take place in well controlled environments. Therefore, testing these processes on real data is needed to confirm the magnitude of these results and the potential for improvement using the choice-based RM concepts.

Chapter 6. SUMMARY, FINDINGS AND RECOMMENDATIONS

When a subject is highly controversial... one cannot hope to tell the truth. One can only show how one came to hold whatever opinion one does hold. One can only give one's audience the chance of drawing their own conclusions as they observe the limitations, the prejudices, the idiosyncrasies of the speaker.

Virginia Woolf, English novelist and essayist
(www.thinkexist.com)

6.1. Summary

This study investigates the revenue performance of choice-based revenue management (RM) systems in various business environments. Previous research conducted using simulated data suggests that incremental revenue gains of up to 15% are to be expected when choice-based RM techniques are employed. In addition, despite the novelty of these techniques, the implementation of choice-based RM systems is considered to be feasible at large global corporations. The revenue potential and the ease of execution associated with the choice-based methods are examined in the context of a large hotel chain. Customer-centric data which includes transaction and time of booking availability information is collected for five hotel properties located in the continental US. The customer preference for hotel products and their attributes is determined using discrete choice and other ad hoc models of demand. The incompleteness of customer-centric data calls for novel approaches to estimating the discrete choice models. Hence, the expectation-maximization (EM) algorithm that builds on the maximum likelihood estimation technique is employed to account for the missing data. The ad hoc models of

demand, which are intended to replace the computationally involved EM method in certain contexts, command the use of advanced forecasting methods. Optimization techniques that account for the customer purchasing behavior are employed to compute the capacity control policies the hotel operator should follow to maximize its revenues. Results indicate that collecting customer-centric data from today's RM systems is a time-consuming task. In the environment in which the study hotels operate, the choice-based RM systems report incremental revenue gains that are dependent on how the purchasing behavior models are formulated. In capacity constrained regimes that are the focus of RM, revenue gains of up to 2% are typically noted. In controlled environments in which the customer purchasing behavior can be better asserted, the incremental revenue gains range between 1% and 14%. In these contexts, the revenue gains are shown to be consistent with respect to how the behavioral inputs are computed. As a result, the ad hoc models appear to be a realistic alternative to the computationally expensive EM procedure. These findings suggest that the execution of the choice-based RM, while feasible, needs to be preceded by the implementation of efficient and, most likely, expensive data collection procedures. The incremental revenue gains, consistent with those reported in the literature, indicate that RM users can substantially benefit from the use of the choice-based RM.

6.2. Contributions and Findings

Within the choice-based RM literature there are two areas of practical contribution of this study. The first relates to situations that precede a potential implementation of a choice-based RM system. In these cases, interested organizations need to understand the data requirements that such an initiative commands and should plan for data collection systems and procedures that support its execution. The second refers to theoretical aspects that current choice-based RM methodologies appear to overlook. In particular, while they emphasize the importance of integrating the customer purchasing behavior

into the standard optimization routines and suggest techniques to estimate the corresponding behavioral models, these methodologies employ simplistic model specifications that are not truly representative of how customers behave in the market place. In addition, the formulation and the estimation of these models lead to significant changes in the results reported when small adjustments to model specifications are undertaken. In this context, we provide guidelines on how the model specification impacts all downstream RM processes including the overall revenue performance.

Compiling customer-centric data has shown that collecting this type of data, post-processing it and checking for its validity is a daunting and time-consuming task. Despite the small number of hotel properties for which we have attempted to obtain customer-centric data, the amount of effort that was required to accomplish the task is illustrative of the difficulties that service providers would face when implementing a choice-based RM system (see Chapter 4 for more details). The extensive use of advanced statistical techniques and outsourced services such as seeking the professional advice of experts employed in areas other than revenue management has revealed the fact that some firms may not be prepared from an institutional perspective to implement advanced RM systems. Therefore, it is our expectation that unless alternative data collection procedures are devised and successfully tested, some service operators and, in particular, those without highly trained staff will be reluctant to invest into choice-based RM systems simply because they do not have the capabilities to manage the expected data requirements. As a result, this study contributes to advancing our understanding of the choice-based RM methodologies by being the first study that identifies and quantifies the practical obstacles that service providers need to address prior to the implementation of advanced RM systems.

From a theoretical perspective, this study contributes to a better understanding of how the specification of customer purchasing behavior models impacts the recommended capacity control policies and the revenue performance of the corresponding choice-based

RM systems. Our extensive testing on real data sets has shown that behavioral interpretations of the results returned by the Expectation - Maximization (EM) algorithm change from one model specification to another. In this context, the counter-intuitive estimates and the mixed result interpretations can be attributed to the formulation and estimation of the behavioral models. The instability of behavioral models' parameter estimates impacts all downstream RM processes and generally leads to recommended capacity control policies that, while theoretically sound, are not completely representative of the markets they are supposed to serve. In such cases, the choice-based control policies tend to marginally outperform the traditional control policies used to manage a company's perishable products. The study of business environments in which customers purchase on price shows that when the customer purchasing behavior is understood the use of choice-based RM techniques leads to consistent and significant revenue benefits. In these environments, a theoretical alternative to the computationally intensive EM algorithm is proposed. This procedure provides for intuitive and simple means to account for the customer purchasing behavior while computing the recommended capacity control policies. By studying the behavioral models in the choice-based RM framework, we have learned that the RM field has not yet exploited the power of these models to its full potential. We believe that behavioral models are currently used as tools for providing a structured way to assigning and redistributing sets of purchasing probabilities required by the control optimizers rather than instruments that soundly describe the customer purchasing behavior. To accomplish a full integration of the customer behavior into the RM practices, it is our expectation that more advanced and/or alternative estimation and optimization techniques need to be devised.

Based on the experience gained from developing, testing, and implementing choice-based RM solutions, a summary of the findings of this research follows next:

- Compiling customer-centric data from the current RM systems turns out to be an involved task. The effort in terms of time, knowledge, system and staff resources

required to collect such data may deter potential users from shifting to choice-based RM systems. Streamlined data collection methodologies need to be made available to warrant the success of the choice-based RM (Section 1.2, Objective 1).

- From an optimization stand point, the choice-based RM methodologies are robust. However, due to the assumptions that govern the estimation of the behavioral models, choice-based methodologies do not fully exploit the power of these models. In particular, the impact of proposed estimation techniques on the behavioral outputs appears to not be well understood. In this context, refinements in model formulation may lead to changes in the underlying customer purchasing behavior (Section 1.2, Objective 2).
- In various contexts, the techniques used to estimate the behavioral models are sensitive to how these models are specified. This impacts all downstream RM processes and generally leads to recommended capacity control policies that, while theoretically sound, are not representative of the markets they are supposed to serve. In such cases, the choice-based control policies tend to marginally outperform the traditional control policies used to manage a company's perishable products (Section 1.2, Objective 3).
- In the environment in which the study hotels operate, the choice-based RM systems report incremental revenue gains that are dependent on how the purchasing behavior models are formulated. In capacity constrained regimes, revenue gains of up to 2% are typically noted. In controlled environments in which the customer purchasing behavior can be better asserted, the incremental revenue gains range between 1% and 14% (Section 1.2, Objective 4).

6.3. Limitations and Future Research

In the last years, the choice-based RM has emerged as a credible alternative to the traditional product-based RM. In this study, we have investigated some of the problems

that other researchers have overlooked and that may impact the implementation and acceptance of these promising concepts. Our findings, while encouraging, should be carefully interpreted in connection with the study limitations. A list of these limitations is given in Table 25. Wherever appropriate, considerations on how future research can lessen the impact of these limitations are provided. At a different level, specific directions for further research are outlined in the next paragraphs. Specifically, recommendations that address the major issues that may deter service providers from shifting to choice-based RM are offered.

The data collection methodology described in this study, while comprehensive, does not support the simultaneous execution of choice-based RM at all hotels affiliated with the hotel chain. Its time-of-booking availability component, which consumes substantial system resources, is too slow and inefficient for the entire procedure to appeal to interested users. Therefore, if the choice-based RM were indeed desired, adequate data collection tools would have to be devised prior to any attempt to operationalize the choice-based concepts. Web-hosted applications able to communicate with the CRS in real time and collect point of sale availability data may provide service operators on legacy mainframe systems with the required flexibility. Alternatively, if a revenue systems transformation is envisioned, the collection of customer-centric data need to be proposed as one of the core capabilities of the new system. In both cases, further research is required to ensure that the devised data collection tools are not obsolete at the time of their release due to advances in technology.

The integration of behavioral models into the optimization routines that lead to the recommended control policies poses significant challenges. During the course of this study, we noticed that researchers focused on finding estimation methods suitable for the characteristics of the customer-centric data with little being done to assert the extent to which the computed behavioral models are representative of the customer population that they are intended to represent. From our experience in developing choice-based RM

solutions we believe that future research is needed to (1) better understand how model formulation impacts the underlying customer purchasing behavior, (2) better comprehend how the estimation algorithms compute the corresponding choice models, and (3) better apprehend how discrete choice models interact with the allocation algorithms in realistic business contexts.

The findings of the simulation study that intended to analyze the choice-based RM in a pure price sensitive environment, while encouraging, should be taken with caution. The assumptions that control the selling of the products in this artificial environment may not be representative of how customers purchase in a real world context. Thus, to validate the magnitude of these results, further research on extending the scope of the exercise to incorporate the actual customer purchasing behavior is needed. To accomplish this, however, a new data collection component needs to be devised and integrated into the overall data collection framework we proposed. In particular, besides the availability data, information about the capacity controls in place at the time of booking needs to be collected. The transaction data together with the time of booking availability and control information will allow us to repeat the simulation experiment in real business environments that are price sensitive.

Table 25 Study Limitations

LIMITATIONS	
1. Single Leg RM Problem (vs. Network RM Problem)	The proposed methodology attempts to manage the hotel capacity as if the demand for each check-in date were time independent. Specifically, given the added complexity, we do not explicitly consider here the case when customers stay multiple hotel nights and the capacity on any given check-in date is shared by customers with different lengths of stay. Instead, we assume that a multiple night stay is equivalent with multiple one night stays and manage the capacity accordingly.
2. Cancellations, No-shows, and Group Arrivals	Given that the focus of this research is on how to effectively incorporate customer purchasing behavior in the hotel RM framework, we do not attempt to include cancellations, no-shows, and group arrivals in our study. Instead, we let these issues be part of future research.
3. Choice/Consideration Set Recreation	The recreation of time-of-booking choice/consideration sets is a difficult, if not impossible, task. Throughout this study, all products available to the customer at the time of booking entered his choice/consideration set. Real time, point-of sale data collection methods intended to record the products a customer looks at may provide researchers with unbiased such information.
3. MNL Model	The MNL model represents a facile way to assign purchasing probabilities to various products in a consideration set. Due to its IIA property, once estimated on a set of independent products, the MNL model can be applied to any incomplete consideration set to estimate these purchasing probabilities. This approach, as simple as it is, may in fact lead to biased results especially when customers express strong preferences for groups of products. Given that there are no documented ways to correct for the potential bias, we stay with the MNL model but acknowledge its limitations.
4. Incomplete Specification of the MNL Utility Function	As pointed out throughout the study, customer choice is assumed to be influenced by the product characteristics only. Considering that customers are most of the time reluctant to disclose their socio-demographic characteristics, this assumption does not appear to notably limit the applicability of the current approach. However, if customer characteristics were available and incorporated in the choice model, then, during the estimation of the optimal choice-based control policy we could run into problems since we had to assign such characteristics to any generic arrival.
5. Fixed Prices for Determining the Choice-Based Control Policy	Room prices can, generally, vary during the booking horizon. The price variability together with the customer price sensitivity is appropriately reflected in the choice model obtained from the hotel booking data. However, given that little is known about how hotels update their prices, we are forced to use fixed, arbitrarily set prices when we determine the optimal choice-based capacity control policy. To control for the potential bias, we use an appropriately adjusted price as the input for the policy optimizer.
3. Distribution Channels	The results reported on real hotel data do not take into account the impact that the distribution channels have on customers' choice. The limited customer-centric data available at the distribution channel level made the estimation of behavioral models unattainable at this level of aggregation. Since customers' price sensitivity has been consistently shown to be influenced by this attribute, future work on the choice-based RM must account for the source of booking when estimating the customer purchasing behavior.
6. Strategic Customer Behavior Models	When informed (e.g., Thanksgiving is well known for the steep discounts most of the retail stores offer during that period; Macy's informs its clientele about the coming discounts two weeks in advance of the sale event), customers may postpone their purchasing decisions in an attempt to take advantage of the expected future low prices. Our current methodological approach cannot describe this strategic game between the firm and its customers, that is, we cannot determine the most likely strategy that customers will adopt in response to the selling strategy adopted by the firm.

Appendix A. EMSR-b REVENUE PERFORMANCE EVALUATION - ALGORITHM DESCRIPTION

Input Files

File Name	Description
Prices.csv	<p>Provides the main procedure with the product class (PC) prices. Its first line, preceded by a # character, gives the user the order in which the prices should be specified. The second line consists of a list of prices delimited by commas.</p> <p>Number of Records: 2 Number of Fields: Number of Product Classes</p> <p>Example: # Prices for: PC1, PC2, PC3, PC4, PC5, PC6, PC7, PC8, PC9 260,280,370,280,310,400,300,340,430</p>
Demand.csv	<p>Provides the main procedure with the latent product class demand at the booking interval level.</p> <p>Number of Records: (Number of Product Classes x Number of Booking Intervals) + 1 Number of Fields: 5 (i.e., Product Class, Booking Interval, Weekdays (0/1), Average Demand, Demand Standard Deviation)</p> <p>Example: PC, Booking Interval, Weekdays, Average Demand, Demand St. Dev. ... PC5,7,0,0.0,0.00 PC5,6,0,4.1,2.02 PC5,5,0,4.8,2.10 PC5,4,0,7.0,4.55 PC5,3,0,7.0,4.00 PC5,2,0,3.4,3.17 PC5,1,0,3.2,3.71 ...</p>

Output Files

File Name	Description
Results.csv	<p>For each of the Phase 2 arrival streams, provides revenue performance statistics (EMSR-b, dynamic update, with or without buy-up). In addition, revenue summary statistics across all arrival streams are given.</p> <p>Number of Records: Varies Number of Fields: NA</p> <p>Example: Arrival Stream: 125 Revenue: 25670 Unsold Capacity: 4 units Accepted Requests (by Product Class): PC2 PC4 PC5 PC6 PC7 PC8 PC9 9 1 18 12 5 26 5 Rate Denials (by Product Class): PC1 PC2 PC4 PC5 6 1 2 2 House Denials: No House Denials</p> <p>Average Revenue over 125 Arrival Streams: 25990 Standard Deviation over 125 Arrival Streams: 1305.81</p>
Results_EMSRb_Static.csv	<p>For each of the Phase 2 arrival streams, provides revenue performance statistics (EMSR-b, static, with or without buy-up). In addition, revenue summary statistics across all arrival streams are given.</p> <p>Number of Records: Varies Number of Fields: NA</p> <p>Example: Arrival Stream: 125 Revenue: 20650 Unsold Capacity: 20 units Accepted Requests: PC2 PC4 PC5 PC6 PC7 PC8 PC9 7 1 12 12 2 21 5 Rate Denials: PC1 PC2 PC4 PC5 6 1 2 2 House Denials: No House Denials</p> <p>Average Revenue over 125 Arrival Streams: 22079.2 Standard Deviation over 125 Arrival Streams: 1882.8</p>
ArrivalStreams.csv	<p>For each of the Phase 2 arrival streams, provides the order of product class requests by Booking Interval.</p> <p>Number of Records: (Number of Booking Intervals + 1) per Arrival Stream Number of Fields: NA</p> <p>Example: Arrival Stream: 1 PC1, PC2, PC6, ... PC7, PC4, PC6, ... PC2, PC6, PC3, ... PC6, PC8, PC8, ... PC9, PC9, PC8, ... PC6, PC5, PC5, ... PC6, PC5, PC8, ...</p>

Procedure EMSRb_BuyUp.R

Input Parameters:

Fare	Numeric vector of product class prices (ordered in ascending order).
Mean	Numeric vector of product class average demands (by booking horizon or booking interval).
StdDev	Numeric vector of standard deviations for the product class demands (by booking horizon or booking interval).
p_up	Numeric vector of buy-up probabilities. If zero, the standard EMSR-b method is employed.
cap	Remaining capacity (numeric).

Output Parameters:

p	Numeric vector of product class protection levels.
---	--

Pseudocode:

```
Procedure EMSRb_BuyUp.R
  Determine the number of product classes
  Compute the variance of the product class demands
  Initialize the product class protection level vector
  FOR all product classes
    Compute the weighted-average revenue for all higher product
      classes
    Compute the aggregate demand for all higher product classes
    Compute the standard deviation for the aggregate demand
    Compute the protection level for current and all higher product
      classes
  Set the last element of the protection level vector to remaining capacity
  Set all protection levels greater than the remaining capacity to the
    remaining capacity
  Round the protection levels to the smallest integers greater than their
    values
  Return the vector of protection levels
```

Procedure Process_Demand.R

Input Parameters:

demand_bi	Numeric vector that provides the expected demands by product class and booking interval.
prot_l	Numeric vector of protection levels to be used at the beginning of the current booking interval.
Cap	Remaining capacity (numeric value).
revenue	Numeric vector that consists of the prices of accepted product classes.
accept	Numeric vector that consists of all accepted requests. The names associated with each vector element identify the accepted product classes.
regret	Numeric vector that consists of all rate denied requests. The names associated with each vector element identify the rate denied product classes.
deny	Numeric vector that consists of all house denied requests. The names associated with each vector element identify the house denied product classes.
arrivalstream	Numeric vector that stores all arrivals for an arrival stream.

Output Parameters:

Cap	Updated remaining capacity (numeric value).
revenue	Updated revenue vector.
accept	Updated vector of accepted requests.
regret	Updated vector of rate denied requests.
deny	Updated vector of house denied requests.
arrivalstream	Updated vector of requests.

Pseudocode:

Procedure Process_Demand.R

```
IF there is demand during the current booking interval THEN
    Determine the product classes for which demand exists
    Compute the corresponding vector of 1 unit requests
    Assign the 1 unit requests a random order of arrival
    Write to ArrivalStreams.csv
    Append the 1 unit requests to arrivalstream vector
    FOR all requests in the 1 unit request vector
        Determine the product class associated with the request
        IF capacity is sold out THEN
            Append the request to deny vector
        ELSE
            IF the request belongs to the highest prod class THEN
                Append the request to the accept vector
                Append the revenue to the revenue vector
                Update capacity to reflect the sale
            ELSE
                IF we can still accept the request THEN
                    Append the request to the accept vector
                    Append the revenue to the revenue
                    vector
                    Update capacity to reflect the sale
                ELSE
                    Append the request to the regret vector
                END IF
            END IF
        END IF
    END IF
ELSE
    Print No Demand
END IF
Return all output parameters
```

Procedure Process_Demand_EMSRb_Static.R

Input Parameters:

demand_bi	Numeric vector that provides the expected demands by product class and booking interval.
prot_l	Numeric vector of protection levels to be used at the beginning of the current booking interval.
Cap	Remaining capacity (numeric value).
revenue	Numeric vector that consists of the prices of accepted product classes.
accept	Numeric vector that consists of all accepted requests. The names associated with each vector element identify the accepted product classes.
regret	Numeric vector that consists of all rate denied requests. The names associated with each vector element identify the rate denied product classes.
deny	Numeric vector that consists of all house denied requests. The names associated with each vector element identify the house denied product classes.

Output Parameters:

Cap	Updated remaining capacity (numeric value).
revenue	Updated revenue vector.
accept	Updated vector of accepted requests.
regret	Updated vector of rate denied requests.
deny	Updated vector of house denied requests.

Pseudocode:

```
Procedure Process_Demand.R
  IF there is demand during the current booking interval THEN
    Compute the corresponding vector of 1 unit requests based on
      demand_bi
    FOR all requests in the 1 unit request vector
      Determine the product class associated with the request
      IF capacity is sold out THEN
        Append the request to deny vector
      ELSE
        IF the request belongs to the highest prod class THEN
          Append the request to the accept vector
          Append the revenue to the revenue vector
          Update capacity to reflect the sale
        ELSE
          IF we can still accept the request THEN
            Append the request to the accept vector
            Append the revenue to the revenue
              vector
            Update capacity to reflect the sale
          ELSE
            Append the request to the regret vector
          END IF
        END IF
      END IF
    END IF
  ELSE
    Print No Demand
  END IF
  Return all output parameters
```

Main Procedure 20070904_EMSRb_Simulation.R

Input Parameters: NA

Output Parameters: NA

Pseudocode:

```
Main Procedure 20070904_EMSRb_Simulation.R
  Read prices from Prices.csv
  Get the number of product classes
  Read latent demand from Demand.csv
  Get the number of booking intervals
  Populate latent demand list by product class and booking interval
  Specify the number of arrival streams associated with Phase 1
  Specify the number of arrival streams associated with Phase 2
  Specify the available capacity
  Specify the booking interval at the beginning of which some of the
    product classes will not be available anymore
  ##### PHASE 1 - BEGIN #####
  Initialize realized demand list by product class and booking interval
  FOR all arrival streams associated with Phase 1
    FOR all product classes
      Randomly sample from the latent demand distributions the
        demands for all booking intervals
      Round these demands to the nearest integers if positive, to
        zero otherwise
    FOR all booking intervals
      Populate the realized demand list by product class
        and booking intervals with the expected demand
        from the beginning of the booking interval
        until the end of the booking horizon
  ##### PHASE 1 - END #####
  ##### PHASE 2 - BEGIN #####
  Initialize the output files
  Initialize the input parameters for EMSRb_BuyUp.R procedure
  Initialize dynamic/static global revenue performance measures
  FOR all arrival streams associated with Phase 2
    ##### DYNAMIC PROTECTION LEVEL UPDATE #####
    Initialize dynamic local revenue performance measures
    Initialize temporary demand structure to facilitate the update of
      the realized demand list
    Print to ArrivalStreams.csv
    FOR booking intervals for which all prod. classes are available
      FOR all product classes
        Determine the demands until the end of the booking
          horizon using the realized demand list
        Determine protection levels: CALL EMSRb_BuyUp.R
        Generate demand for current booking interval using the
          latent demand distributions
        Populate temporary demand structure
        Process the booking interval demand: CALL Process_Demand.R
        Update dynamic local revenue performance measures
    FOR booking intervals for which some prod. classes are available
      FOR all available product classes
        Determine the demands until the end of the booking
          horizon using the realized demand list
        Determine protection levels: CALL EMSRb_BuyUp.R
        Generate demand for current booking interval using the
          latent demand distributions
        Populate temporary demand structure
        Process the booking interval demand: CALL Process_Demand.R
```



```

        Update dynamic local revenue performance measures
##### STATIC PROTECTION LEVELS #####
Initialize static local revenue performance measures
FOR all product classes
    Determine the demands for the entire booking horizon using
        the realized demand list
    Determine protection levels: CALL EMSRb_BuyUp.R
    Process the entire booking horizon demand: CALL
        Process_Demand_EMSRb_Static.R
    Update static local revenue performance measures
##### PRINT RESULTS #####
Print static local revenue performance measures to
    Results_EMSRb_Static.csv
Update realized demand list
Print dynamic local revenue performance measures to Results.csv
Update dynamic/static global revenue performance measures
Print static global revenue performance measures to
    Results_EMSRb_Static.csv
Print dynamic global revenue performance measures to Results.csv
##### PHASE 2 - END #####

```

REFERENCES

- [1] SAS Institute Inc., *Statistical Analysis System: Release 9.1*. Cary, NC: SAS Institute Inc., 2007.
- [2] R. G. Cross, *Revenue Management: Hard-Core Tactics for Market Domination*. New-York: Broadway Books, 1997.
- [3] "Airline Deregulation Act," in *Public Law no. 95-504, 92 Stat. 1750*, 1978.
- [4] E. A. Boyd, "Dramatic Changes in Distribution will Require Renewed Focus on Pricing and Revenue Management Models," *Journal of Revenue & Pricing Management*, vol. 3, pp. 100-103, 2004.
- [5] E. A. Boyd and R. Kallesen, "The Science of Revenue Management when Passengers Purchase the Lowest Available Fare," *Journal of Revenue & Pricing Management*, vol. 3, pp. 171-177, 2004.
- [6] H. Dunleavy and D. Westermann, "Future of Airline Revenue Management," *Journal of Revenue & Pricing Management*, vol. 3, pp. 380-383, 2005.
- [7] P. Hoang, "The Future of Revenue Management and Pricing Science," *Journal of Revenue & Pricing Management*, vol. 6, pp. 151-153, 2007.
- [8] S. Hornick, "Fare Play," *Airline Business*, vol. 20, pp. 72-75, 2004.
- [9] W. H. Lieberman, "Revenue Management Trends and Opportunities," *Journal of Revenue & Pricing Management*, vol. 3, pp. 91-99, 2004.
- [10] S. D. Nason, "Forecasting the Future of Airline Revenue Management," *Journal of Revenue & Pricing Management*, vol. 6, pp. 64-66, 2007.
- [11] V. M. Oliveira, "Simulating Revenue Management in an Airline Market with Demand Segmentation and Strategic Interaction," *Journal of Revenue & Pricing Management*, vol. 1, pp. 301, 2003.
- [12] R. Ratliff, "Ideas on Overbooking," presented at AGIFORS Reservations and Yield Management Study Group, Melbourne, Australia, 1998.
- [13] F. Sfodera, *The Spread of Yield Management Practices: The need for Systematic Approaches*. Heidelberg, Germany: Physica-Verlag, 2006.

- [14] K. T. Talluri and G. van Ryzin, "Revenue Management under a General Discrete Choice Model of Consumer Behavior.," *Management Science*, vol. 50, pp. 15-33, 2004.
- [15] G. van Ryzin, "Models of Demand.," *Journal of Revenue & Pricing Management*, vol. 4, pp. 204-210, 2005.
- [16] D. Zhang and W. Cooper, "Revenue Management for Parallel Flights with Customer-Choice Behavior," *Operations Research*, vol. 53, pp. 415-431, 2005.
- [17] G. Gallego, G. Iyengar, R. Phillips, and A. Dubey, "Managing Flexible Products on a Network," Department of Industrial Engineering and Operations Research, Columbia University, CORC Technical Report TR-2004-01, 2004.
- [18] G. van Ryzin and Q. Liu, "On the Choice-Based Linear Programming Model for Network Revenue Management," Columbia Business School, Decision, Risk & Operations Working Papers Series, DRO-2004-04, 2004, http://www2.gsb.columbia.edu/divisions/dro/working_papers2004.html, Access Date: January 2007
- [19] G. van Ryzin and G. Vulcano, "Computing Virtual Nesting Controls for Network Revenue Management under Customer Choice Behavior," Columbia Business School, Decision, Risk & Operations Working Paper Series, DRO-2004-09, 2004, http://www2.gsb.columbia.edu/divisions/dro/working_papers2004.html, Access Date: January 2007
- [20] G. van Ryzin and G. Vulcano, "Simulation-Based Optimization of Virtual Nesting Controls for Network Revenue Management," *Forthcoming in Operations Research*, 2006.
- [21] J. J. M. Bront, I. Mendez-Diaz, and G. Vulcano, "A Column Generation Algorithm for Choice-Based Network Management," pp. 25, Paper Submitted to AGIFORS for Anna Valicek Medal 2007, 2007.
- [22] G. Gallego and O. Sahin, "Inter-Temporal Valuations, Product Design and Revenue Management," pp. Paper Submitted to AGIFORS for Anna Valicek Medal 2006 (Runner-Up), 2006.
- [23] M. Hu and G. Gallego, "Dynamic Pricing of Perishable Assets under Competition," pp. Paper Submitted to AGIFORS for Anna Valicek Medal 2007, 2007.

- [24] American Airlines Inc., "Annual Report. The Art of Managing Yield," American Airlines Inc. 1987.
- [25] B. Smith, J. Leimkuhler, and R. Darrow, "Yield Management at American Airlines," *Interfaces*, vol. 22, pp. 8-31, 1992.
- [26] R. D. Hanks, R. G. Cross, and R. P. Noland, "Discounting in the Hotel Industry: A New Approach," *Cornell Hotel and Restaurant Administration Quarterly*, vol. 43, pp. 94-103, 2002.
- [27] S. E. Kimes, "The Basics of Yield Management.," *Cornell Hotel & Restaurant Administration Quarterly*, vol. 30, pp. 14, 1989.
- [28] E. B. Orkin, "Strategies For: Managing Transient Rates," *Cornell Hotel and Restaurant Administration Quarterly*, vol. 30, pp. 35, 1990.
- [29] B. Vinod, "Unlocking the value of revenue management in the hotel industry.," *Journal of Revenue & Pricing Management*, vol. 3, pp. 178-190, 2004.
- [30] T. Matanovich, G. L. Lilien, and A. Rangaswamy, "Engineering the Price-Value Relationship," *Marketing Management*, vol. 8, pp. 48-53, 1999.
- [31] R. C. Lewis and S. Shoemaker, "Price-Sensitivity Measurement : A Tool for the Hospitality Industry," *The Cornell Hotel and Restaurant Administration Quarterly*, vol. 38, pp. 44-54, 1997.
- [32] E. Steed and Z. Gu, "An Examination of Hotel Room Pricing Methods: Practised and Proposed.," *Journal of Revenue & Pricing Management*, vol. 3, pp. 369-379, 2005.
- [33] K. Varini, R. Engelmann, B. Claessen, and M. Schleusener, "Evaluation of the Price-Value Perception of Customers in Swiss Hotels.," *Journal of Revenue & Pricing Management*, vol. 2, pp. 47, 2003.
- [34] A. O. Lee, "Airline Reservations Forecasting: Probabilistic and Statistical Models of the Booking Process," in *Department of Civil and Environmental Engineering*, vol. Ph.D. Cambridge, Massachusetts: Massachusetts Institute of Technology, 1990.
- [35] K. T. Talluri and G. van Ryzin, *The Theory and Practice of Revenue Management*. New York, NY: Kluwer Academic Press, 2004.

- [36] P. P. Belobaba, "Airline Yield Management: An Overview of Seat Inventory Control," *Transportation Science*, vol. 21, pp. 63-73, 1987.
- [37] S. L. Brumelle, J. McGill, T. H. S. Oum, K., and M. W. Tretheway, "Allocation of Airline Seats Between Stochastically Dependent Demands," *Transportation Science*, vol. 24, pp. 183-192, 1990.
- [38] P. E. Pfeifer, "The Airline Discount Fare Allocation Problem," *Decision Sciences*, vol. 20, pp. 149, 1989.
- [39] P. P. Belobaba, "Air Travel Demand and Airline Seat Inventory Management," in *Flight Transportation Laboratory*, vol. Ph.D. Thesis. Cambridge, MA: Massachusetts Institute of Technology, 1987.
- [40] P. P. Belobaba and L. R. Weatherford, "Comparing Decision Rules that Incorporate Customer Diversion in Perishable Asset Revenue Management Situations," *Decision Sciences*, vol. 27, pp. 343, 1996.
- [41] P. P. Belobaba and C. Hopperstad, "Boeing/MIT Simulation Study: PODS Results Update," presented at AGIFORS Reservations and Yield Management Study Group Symposium Proceedings, London, UK, 1999.
- [42] P. P. Belobaba and J. L. Wilson, "Impacts of Yield Management in Competitive Airline Markets," *Journal of Air Transport Management*, vol. 3, pp. 3-9, 1997.
- [43] P. P. Belobaba, "Revenue and Competitive Impacts of O-D Control: Summary of PODS Results," presented at INFORMS Revenue Management Conference, New-York, USA, 2001.
- [44] E. Carrier, "Modeling Airline Passenger Choice: Passenger Preference for Schedule in the Passenger Origin-Destination Simulator (PODS)," in *Department of Civil and Environmental Engineering*, vol. M.S. Thesis. Cambridge, MA: Massachusetts Institute of Technology, 2003.
- [45] G. Vulcano, G. van Ryzin, and W. Chahr, "Choice-Based Revenue Management: An Empirical Study of Estimation and Optimization," presented at INFORMS 2006, Pittsburgh, PA, 2006.
- [46] C. H. Bohutinsky, "The Sell-Up Potential of Airline Demand," in *Flight Transportation Laboratory*, vol. M.S. Thesis. Cambridge, MA: Massachusetts Institute of Technology, 1990.

- [47] M. Rothstein, "Stochastic Models for Airline Booking Policies," in *Graduate School of Engineering and Science*, vol. Ph.D. Thesis. New York: New York University, 1968.
- [48] E. L. Williamson, "Airline Network Seat Inventory Control: Methodologies and Revenue Impacts," in *Flight Transportation Laboratory*, vol. Ph.D. Thesis. Cambridge, MA: Massachusetts Institute of Technology, 1992.
- [49] R. E. Curry, "Optimal Airline Seat Allocation with Fare Classes Nested by Origins and Destinations," *Transportation Science*, vol. 24, pp. 193-204, 1990.
- [50] R. D. Wollmer, "An Airline Seat Management Model for a Single Leg Route when Lower Fare Classes Book First," *Operations Research*, vol. 40, pp. 26, 1992.
- [51] S. L. Brumelle and J. I. McGill, "Airline Seat Allocation with Multiple Nested Fare Classes," *Operations Research*, vol. 41, pp. 127, 1993.
- [52] L. W. Robinson, "Optimal and Approximate Control Policies for Airline Booking with Sequential Nonmonotonic Fare," *Operations Research*, vol. 43, pp. 252, 1995.
- [53] T. C. Lee and M. Hersh, "A Model for Dynamic Airline Seat Inventory Control with Multiple Seat Bookings," *Transportation Science*, vol. 27, pp. 252-265, 1993.
- [54] C. J. Lautenbacher and S. J. Stidham, "The Underlying Markov Decision Process in the Single-Leg Airline Yield Management Problem," *Transportation Science*, vol. 34, pp. 136-146, 1999.
- [55] C. Hopperstad, "Potential Application of the Boeing Passenger Origin-Destination Simulator (PODS) to Yield Management System Planning.," Revision C. Boeing Commercial Airplane Group, February 20, 1995.
- [56] M. D. Meyer and E. J. Miller, *Urban Transportation Planning: A Decision-Oriented Approach*, Second ed. Boston: McGraw-Hill Higher Education 2001.
- [57] W. L. Winston and M. Venkataramanan, *Introduction to Mathematical Programming*, Fourth ed: Duxbury Press, 2003.
- [58] D. McFadden, "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, P. Zarembka, Ed. New York: Academic Press, 1974, pp. 75-96.

- [59] D. McFadden, "Modeling the Choice of Residential Location," in *Spatial Interaction Theory and Planning Methods*, A. Karlquist, L. Lundqvist, F. Snickars, and J. W. Weibull, Eds. Amsterdam: North Holland, 1978, pp. 75-96.
- [60] H. C. W. L. Williamson, "On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit," *Environment and Planning A*, vol. 9(3), pp. 285-344, 1977.
- [61] C. Daganzo, *Multinomial Probit: The Theory and Its Application to Demand Forecasting*. New York: Academic Press, 1979.
- [62] J. Marschak, "Binary Choice Constraints on Random Utility Indications," in *Mathematical Methods in the Social Sciences*, K. Arrow, Karlin, and Suppes, Eds. Stanford, CA.: Stanford University Press, 1960, pp. 312-329.
- [63] L. L. Thurstone, "A Law of Comparative Judgement," *Psychological Review*, vol. 34, pp. 273-286, 1927.
- [64] O. Petrova, "Group Discussion about Internet Marketing Strategies," T. Bodea, Ed. Atlanta, 2007.
- [65] B. Smith, J. Elieson, R. Darrow, D. Guenther, F. Zouaoui, and B. Rao, "Retail Modeling for Travelocity," presented at AGIFORS Annual Symposium 2006, Rome, Italy, 2006.
- [66] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, pp. 1-38, 1977.
- [67] P. P. Belobaba, "Application of a Probabilistic Decision Model to Airline Seat Inventory Control," *Operations Research*, vol. 37, pp. 183-197, 1989.
- [68] K. Littlewood, "Forecasting and Control of Passenger Bookings," presented at AGIFORS Annual Symposium 1972, Nathanya, Israel, 1972.
- [69] S. Nahmias, *Production and Operations Analysis*, 5th ed. Boston: McGraw-Hill Irwin, 2005.
- [70] R Development Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2008.

- [71] T. Bodea, M. Ferguson, and L. Garrow, "Choice-based Revenue Management: Data from a Major Hotel Chain," *Manuscript Submitted to Manufacturing and Service Operations Management*, 2007.
- [72] Microsoft Corporation, *Visual Basic*: Microsoft Corporation, 2007.
- [73] ActiveState Software Inc., *ActivePerl: Release 5.8.8.820*: ActiveState Software Inc., 2007.
- [74] StataCorp., *Stata Statistical Software: Release 9.2*. College Station, TX: StataCorp LP, 2007.
- [75] T. Bodea, M. Ferguson, and L. Garrow, "Choice-Based Revenue Management Techniques in Price Sensitive Business Environments," Georgia Institute of Technology, Working Paper, 2008,
- [76] J. Y. Bakos, "Reducing buyer search costs: Implications for electronic marketplaces," *Management Science*, vol. 43, pp. 1676-1692, 1997.
- [77] Doubleclick Performics, "Challenges and Best Practices for Achieving Online Shopper Loyalty: A Look Inside the E-Commerce Consumer Experience," vol. 2008 from www.performics.com, Access Date: February 2008, 2007.
- [78] G. Howlett, "Price Still Most Important Factor for Retail Success, But Losing Ground," vol. 2008 from www.marketingpilgrim.com, Access Date: January 2008, 2007.
- [79] K. C. Jones, "Online Shoppers Care Most About Price," vol. 2008 from <http://informationweek.com>, Access Date: February 2008: eBusinessNews.info, 2007.
- [80] M. Starkov and J. Price, "Building a De-Commoditization Strategy in Hospitality," vol. 2008 from www.hotelmarketing.com, Access Date: January 2008, 2007.
- [81] E. A. Boyd, E. Kambour, and J. Tama, "The Impact of Buy Down on Sell Up, Unconstraining, and Spiral Down," presented at 1st Annual INFORMS Revenue Management Section Conference, New York, NY, US, 2001.
- [82] W. L. Cooper, T. Homem-de-Mello, and A. Kleywegt, "Models of the Spiral-Down Effect in Revenue Management," *Operations Research*, vol. 54, pp. 968-987, 2006.

- [83] R. Kuhlmann, "Why Is Revenue Management Not Working?," *Journal of Revenue and Pricing Management*, vol. 2, pp. 378-387, 2004.
- [84] Aptech Systems Inc., "GAUSS Mathematical and Statistical System," Black Diamond, WA 2008.
- [85] L. R. Weatherford and S. Pölt, "Better Unconstraining of Airline Demand Data in Revenue Management Systems for Improved Forecast Accuracy and Greater Revenues," *Journal of Revenue & Pricing Management* vol. 1, pp. 234-254, 2002.
- [86] R. Zeni, "Improved Forecast Accuracy in Airline Revenue Management by Unconstraining Demand Estimates from Censored Data," in *Management Science*, vol. PhD. Newark, NJ: Rutgers University, 2001.
- [87] S. Pölt, "From Bookings to Demand: The Process of Unconstraining," presented at AGIFORS Reservations and Yield Management Study Group, New York, NY, US, 2000.
- [88] R. Zeni, "Improving Forecast Accuracy by Unconstraining Censored Demand Data," presented at AGIFORS Reservations and Yield Management Study Group, Bangkok, Thailand, 2001.
- [89] C. Crystal, M. Ferguson, J. Higbie, and R. Kapoor, "A Comparison of Unconstraining Methods to Improve Revenue Management Systems," *Forthcoming in Production and Operations Management*, 2007.
- [90] J. S. Armstrong, *Principles of Forecasting: A Handbook for Researchers and Practitioners*: Kluwer Academic Publishers, 2001.
- [91] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, 2nd ed. New York, NY: Springer Science and Business Media Inc., 2002.
- [92] C. Chatfield, *Time-Series Forecasting*: Chapman and Hall/CRC Press, 2000.

VITA

TUDOR DAN BODEA

Tudor Dan Bodea was born June 25, 1977 in Cluj-Napoca, Romania. He grew up in Cluj-Napoca and Corabia, eventually studying computer science at the Emil Racovita National College, a nationally acclaimed, public high school.

Tudor attended the Technical University of Cluj-Napoca from 1995 to 2000, and graduated summa cum laude, with a B.S. in Civil Engineering. In 2000, Tudor entered the graduate program at the Technical University of Cluj-Napoca and received an M.S. in Civil and Environmental Engineering in 2001. Simultaneously, Tudor became a lecturer at the Technical University of Cluj-Napoca where he taught several classes in transportation systems.

In 2002, Tudor entered the Ph.D. program at the Georgia Institute of Technology, Atlanta, GA, and received an M.S. in Civil and Environmental Engineering in 2004. In Atlanta, he became interested in the choice-based revenue management while working under the advice of Dr. Laurie Garrow. In 2007, Tudor was awarded the Jean-Lou Chameau Research Excellence Award for his graduate work in the Transportation Group at the Georgia Institute of Technology. In 2008, Tudor received his doctoral degree in Civil Engineering from the Georgia Institute of Technology.